Protein Sequence Searches - February 2005

All of the sequence databases on ABSS have recently been updated.

- Please note that the curators of the UniProt database have purged some temporary accession numbers from the most recent version of UniProt. These sequences have been assigned new permanent accession numbers. The new UniProt record may not contain the previous temporary accession number.

- If you encounter an accession number from an older search run against UniProt (results file extension **.rup**) that can no longer be found in the database, the permanent record with the new accession number can be found by searching the old accession number in the UniProt Protein Archive database (UniPARC) at:

  http://www.pir.uniprot.org/database/archive.shtml

If you have any questions regarding this information or your results, please contact any STIC searcher.

When submitting sequence search results for scanning into IFW, please include a copy of this attachment to assist any future Examiners or members of the public who may encounter UniProt temporary accession numbers.

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: May 17, 2005, 06:13:14 ; Search time 11.25 Seconds
(without alignments)
76.973 Million cell updates/sec

Title: US-10-725-373-2
Perfect score: 45
Sequence: 1 YLSGADLNL 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_79:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	40	88.9	702	2	A36319	carcinoembryonic a
2	36	80.0	834	2	F82673	hypothetical prote
3	36	80.0	953	2	AH1972	hypothetical prote
4	35	77.8	316	2	C86828	2-dehydro-3-deoxyg
5	34	75.6	506	2	C81704	monooxygenase-rela
6	34	75.6	540	2	E69861	ABC transporter (A
7	33	73.3	404	2	C96549	hypothetical prote
8	33	73.3	491	2	A49179	melanoma antigen h
9	33	73.3	596	2	T23685	hypothetical prote
10	33	73.3	856	2	T10585	serine proteinase
11	32	71.1	105	2	AI1604	weakly B. subtilis
12	32	71.1	108	2	S37313	transcription acti
13	32	71.1	167	2	A30169	hypothetical 19.9K
14	32	71.1	167	2	S29515	lktC protein - Pas
15	32	71.1	176	2	F71064	micrococcal nuclea
16	32	71.1	275	2	PN0511	gastrin-binding pr
17	32	71.1	350	2	T22450	hypothetical prote
18	32	71.1	357	2	AD1062	protein kinase [im
19	32	71.1	377	2	G84806	hypothetical prote
20	32	71.1	382	2	B84297	Htr-like protein [
21	32	71.1	391	2	B69252	3-ketoacyl-CoA thi
22	32	71.1	502	2	S50519	hypothetical prote
23	32	71.1	616	2	E85508	hypothetical prote
24	32	71.1	616	2	E90657	hypothetical prote
25	32	71.1	658	2	T40107	hypothetical 57.9
26	32	71.1	661	1	KFHU13	coagulation factor
27	32	71.1	668	2	A46013	coagulation factor
28	32	71.1	714	2	C65007	probable fatty oxi
29	32	71.1	714	2	A85876	probable enzyme Z3
30	32	71.1	714	2	H91031	probable enzyme [i
31	32	71.1	715	2	AB0805	3-hydroxyacyl-CoA
32	32	71.1	737	2	C70132	hypothetical prote
33	32	71.1	763	2	JC2108	long-chain-fatty-a
34	32	71.1	860	2	C72338	hypothetical prote
35	32	71.1	896	2	H70233	hypothetical prote
36	32	71.1	1473	2	T38791	probable ferredoxi
37	32	71.1	1986	2	S28353	probable polyketid
38	31	68.9	176	2	AH0881	probable exported
39	31	68.9	184	1	S47020	ribosomal protein
40	31	68.9	226	2	AC2825	transcription regu
41	31	68.9	231	2	B97603	hypothetical prote
42	31	68.9	265	2	T14645	hypothetical prote
43	31	68.9	284	2	B83362	probable transcrip
44	31	68.9	295	2	T44541	hypothetical prote
45	31	68.9	308	2	A82391	transcription regu

ALIGNMENTS

RESULT 1
A36319
carcinoembryonic antigen precursor - human
N;Alternate names: CEA; meconium antigen 100
C;Species: Homo sapiens (man)
C;Date: 16-Sep-1992 #sequence_revision 16-Sep-1992 #text_change 09-Jul-2004
C;Accession: A36319; A27773; A31037; A25845; S08106; S31737; A44476; I54224; I59098; A26
R;Schrewe, H.; Thompson, J.; Bona, M.; Hefta, L.J.F.; Maruya, A.; Hassauer, M.; Shively,
Mol. Cell. Biol. 10, 2738-2748, 1990
A;Title: Cloning of the complete gene for carcinoembryonic antigen: analysis of its promo
A;Reference number: A36319; MUID:90258861; PMID:2342461
A;Accession: A36319
A;Molecule type: DNA
A;Residues: 1-702 <SCH>
A;Cross-references: UNIPROT:P06731; GB:M17303; NID:g178676; PIDN:AAB59513.1; PID:g178677
A;Note: the authors show the codons TTA for residue 641-Phe and CAG for residue 646-Thr
R;Beauchemin, N.; Benchimol, S.; Cournoyer, D.; Fuks, A.; Stanners, C.P.
Mol. Cell. Biol. 7, 3221-3230, 1987
A;Title: Isolation and characterization of full-length functional cDNA clones for human c
A;Reference number: A27773; MUID:88038876; PMID:3670312
A;Accession: A27773
A;Molecule type: mRNA
A;Residues: 1-702 <BEA>
A;Cross-references: GB:M29540; NID:g180222; PIDN:AAA51967.1; PID:g180223
R;Barnett, T.; Goebel, S.J.; Nothdurft, M.A.; Elting, J.J.
Genomics 3, 59-66, 1988
A;Title: Carcinoembryonic antigen family: characterization of cDNAs coding for NCA and CE
A;Reference number: A31037; MUID:89122014; PMID:3220478
A;Accession: A31037
A;Molecule type: mRNA
A;Residues: 1-702 <BAR>
A;Cross-references: GB:M29540; NID:g180222; PIDN:AAA51967.1; PID:g180223
A;Note: the authors translated the codon GTG for residue 130 as Leu
R;Oikawa, S.; Nakazato, H.; Kosaki, G.
Biochem. Biophys. Res. Commun. 142, 511-518, 1987
A;Title: Primary structure of human carcinoembryonic antigen (CEA) deduced from cDNA sequ
A;Reference number: A25845; MUID:87128144; PMID:3814146
A;Accession: A25845
A;Molecule type: mRNA
A;Residues: 5-702 <OIK>
A;Cross-references: GB:M15042; NID:g180198; PIDN:AAA51963.1; PID:g180199
R;Oikawa, S.
submitted to the EMBL Data Library, September 1989
A;Reference number: S08106
A;Accession: S08106
A;Molecule type: mRNA
A;Residues: 5-319,321-702 <OI2>
A;Cross-references: EMBL:X16455; NID:g29854; PIDN:CAA34474.1; PID:g825638
R;Barnett, T.
submitted to the EMBL Data Library, September 1991
A;Description: Genomic DNA sequence upstream of the translational start of the carcinoemb
A;Reference number: S31737

A;Accession: S31737
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-141 <BA2>
A;Cross-references: EMBL:X62151
R;Khan, W.N.; Fraengsmyr, L.; Teglund, S.; Israelsson, A.; Bremer, K.; Hammarstroem, S.
Genomics 14, 384-390, 1992
A;Title: Identification of three new genes and estimation of the size of the carcinoembr
A;Reference number: A44476; MUID:93052339; PMID:1427854
A;Accession: A44476
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: DNA
A;Residues: 35-141 <KHA>
R;Willcocks, T.C.; Craig, I.W.
Genomics 8, 492-500, 1990
A;Title: Characterization of the genomic organization of human carcinoembryonic antigen
A;Reference number: I54224; MUID:91139118; PMID:2286372
A;Accession: I54224
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-37 <RES>
A;Cross-references: GB:M60964; NID:g180215; PIDN:AAA51964.1; PID:g180217
R;Zimmermann, W.; Ortlieb, B.; Friedrich, R.; von Kleist, S.
Proc. Natl. Acad. Sci. U.S.A. 84, 2960-2964, 1987
A;Title: Isolation and characterization of cDNA clones encoding the human carcinoembryon
A;Reference number: I59098; MUID:87204247; PMID:3033671
A;Accession: I59098
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 331-702 <RE2>
A;Cross-references: GB:M16234; NID:g180240; PIDN:AAA51972.1; PID:g180241
R;Siepen, D.; Paxton, R.J.; Neumaier, M.; Shively, J.E.; Wagener, C.
Biochem. Biophys. Res. Commun. 147, 212-218, 1987
A;Title: Carcinoembryonic antigen (CEA) and two crossreacting antigens of 165 KD and 105
A;Reference number: A26831; MUID:87326349; PMID:3632664
A;Accession: A26831
A;Molecule type: protein
A;Residues: 35-64 <SIE>
R;Thomas, P.; Toth, C.A.
Biochem. Biophys. Res. Commun. 170, 391-396, 1990
A;Title: Carcinoembryonic antigen binding to Kupffer cells is via a peptide located at t
A;Reference number: A35490; MUID:90321257; PMID:2372297
A;Accession: A35490
A;Molecule type: protein
A;Residues: 'X',140-151,'X',153,'X',155-156 <THO>
A;Note: this is the amino terminal end of a fragment shown to mediate uptake by Kupffer
C;Comment: This heavily glycosylated membane protein of unknown function is a widely use
C;Comment: This protein may be processed at its C-terminus. It is anchored to the membra
C;Genetics:
A;Gene: GDB:CEA
A;Cross-references: GDB:119054; OMIM:114890
A;Map position: 19q13.2-19q13.2
A;Introns: 22/1; 142/1; 235/1; 320/1; 413/1; 498/1; 591/1; 676/1
C;Superfamily: carcinoembryonic antigen; carcinoembryonic antigen precursor amino-termin
C;Keywords: blocked carboxyl end; glycoprotein; lipoprotein; membrane protein; phosphati
F;1-138/Domain: carcinoembryonic antigen precursor amino-terminal homology <CEAN>
F;1-34/Domain: signal sequence #status predicted <SIG>
F;35-678/Product: carcinoembryonic antigen #status predicted <MAT>
F;160-217/Domain: immunoglobulin homology <IMM1>
F;252-301/Domain: immunoglobulin homology <IMM2>
F;338-395/Domain: immunoglobulin homology <IMM3>
F;516-573/Domain: immunoglobulin homology <IMM4>
F;608-657/Domain: immunoglobulin homology <IMM5>
F;679-702/Domain: carboxyl-terminal propeptide #status predicted <CTP>
F;678/Modified site: GPI-anchor ethanolamine amidated carboxyl end (Gly) (in mature form

Query Match 88.9%; Score 40; DB 2; Length 702;
Best Local Similarity 88.9%; Pred. No. 5.6;
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 YLSGADLNL 9
||||||:|||
Db 605 YLSGANLNL 613

RESULT 2
F82673
hypothetical protein XF1508 [imported] - Xylella fastidiosa (strain 9a5c)
C;Species: Xylella fastidiosa
C;Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 09-Jul-2004
C;Accession: F82673
R;anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequenc
Nature 406, 151-157, 2000
A;Title: The genome sequence of the plant pathogen Xylella fastidiosa.
A;Reference number: A82515; MUID:20365717; PMID:10910347
A;Note: for a complete list of authors see reference number A59328 below
A;Accession: F82673
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-834 <SIM>
A;Cross-references: UNIPROT:Q9PD71; GB:AE003980; GB:AE003849; NID:g9106531; PIDN:AAF8431
A;Experimental source: strain 9a5c
R;Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; Al
Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrer, H.
as-Neto, E.; Docena, C.; El-Dorry, H.; Facincani, A.P.; Ferreira, A.J.S.
submitted to GenBank, June 2000
A;Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohme
J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laigre
chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E.
A;Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;
, F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A.
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasaki
A;Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveira
M.; Tsuhako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Za
A;Reference number: A59328
A;Contents: annotation
C;Genetics:
A;Gene: XF1508

Query Match 80.0%; Score 36; DB 2; Length 834;
Best Local Similarity 87.5%; Pred. No. 44;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 YLSGADLN 8
|||| |||
Db 391 YLSGMDLN 398

RESULT 3
AH1972
hypothetical protein alr1331 [imported] - Nostoc sp. (strain PCC 7120)
C;Species: Nostoc sp. PCC 7120
A;Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C;Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Jul-2004
C;Accession: AH1972
R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi,
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.
DNA Res. 8, 205-213, 2001
A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anab
A;Reference number: AB1807; MUID:21595285; PMID:11759840
A;Accession: AH1972
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-953 <KUR>
A;Cross-references: UNIPROT:Q8YX84; GB:BA000019; PIDN:BAB73288.1; PID:g17130678; GSPDB:GN
A;Experimental source: strain PCC 7120
C;Genetics:
A;Gene: alr1331

Query Match 80.0%; Score 36; DB 2; Length 953;
Best Local Similarity 87.5%; Pred. No. 51;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 YLSGADLN 8
|||||||:
Db 828 YLSGADLS 835

RESULT 4
C86828
2-dehydro-3-deoxygluconokinase (EC 2.7.1.45) [imported] - Lactococcus lactis subsp. lact
C;Species: Lactococcus lactis subsp. lactis
C;Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 09-Jul-2004
C;Accession: C86828
R;Bolotin, A.; Wincker, P.; Mauger, S.; Jaillon, O.; Malarme, K.; Weissenbach, J.; Ehrli
Genome Res. 11, 731-753, 2001
A;Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis ss
A;Reference number: A86625; MUID:21235186; PMID:11337471
A;Accession: C86828
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-316 <STO>
A;Cross-references: UNIPROT:Q9CF54; GB:AE005176; PID:g12724636; PIDN:AAK05725.1; GSPDB:G
A;Experimental source: strain IL1403
C;Genetics:
A;Gene: kdgK
C;Superfamily: ribokinase
C;Keywords: phosphotransferase

Query Match 77.8%; Score 35; DB 2; Length 316;
Best Local Similarity 66.7%; Pred. No. 25;
Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 YLSGADLNL 9
||:||:||:
Db 32 YLAGAELNV 40

RESULT 5
C81704
monooxygenase-related protein TC0425 [imported] - Chlamydia muridarum (strain Nigg)
C;Species: Chlamydia muridarum, Chlamydia trachomatis MoPn
C;Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 09-Jul-2004
C;Accession: C81704
R;Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey,
, C.; Dodson, R.; Gwinn, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzberg,
Nucleic Acids Res. 28, 1397-1406, 2000
A;Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39.
A;Reference number: A81500; MUID:20150255; PMID:10684935
A;Accession: C81704
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-506 <TET>
A;Cross-references: UNIPROT:Q9PKP0; GB:AE002309; GB:AE002160; NID:g7190464; PIDN:AAF3928
A;Experimental source: strain Nigg (MoPn)
C;Genetics:
A;Gene: TC0425

Query Match 75.6%; Score 34; DB 2; Length 506;
Best Local Similarity 66.7%; Pred. No. 65;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 YLSGADLNL 9
|||| :||:
Db 290 YLSGVNLNI 298

RESULT 6
E69861
ABC transporter (ATP-binding protein) homolog ykpA - Bacillus subtilis
C;Species: Bacillus subtilis
C;Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 16-Aug-2004
C;Accession: E69861
R;Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berter
C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Cho
A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.
Nature 390, 249-256, 1997
A;Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galler
iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.
Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,
A;Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel,
y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle,
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon,
A;Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seror,
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama,
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K.
A;Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.
A;Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
A;Reference number: A69580; MUID:98044033; PMID:9384377
A;Accession: E69861
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-540 <KUN>
A;Cross-references: UNIPROT:O31716; GB:Z99111; GB:AL009126; NID:g2633699; PIDN:CAB13316.1
A;Experimental source: strain 168
C;Genetics:
A;Gene: ykpA
C;Superfamily: ATP-binding cassette homology
C;Keywords: ATP; nucleotide binding; P-loop
F;17-228/Domain: ATP-binding cassette homology <ABC1>
F;34-41/Region: nucleotide-binding motif A (P-loop)
F;335-512/Domain: ATP-binding cassette homology <ABC2>

Query Match 75.6%; Score 34; DB 2; Length 540;
Best Local Similarity 66.7%; Pred. No. 70;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 YLSGADLNL 9
| |:||||
Db 394 YFEGSDLNL 402

RESULT 7
C96549
hypothetical protein F11M15.3 [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 09-Jul-2004
C;Accession: C96549
R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, I
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A;Reference number: A86141; MUID:21016719; PMID:11130712
A;Accession: C96549
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-404 <STO>
A;Cross-references: UNIPROT:Q9SYB9; GB:AE005173; NID:g4836928; PIDN:AAD30630.1; GSPDB:GN(
C;Genetics:
A;Gene: F11M15.3
A;Map position: 1
C;Superfamily: probable serine/threonine-specific protein kinase ATPK64; protein kinase k

Query Match 73.3%; Score 33; DB 2; Length 404;
Best Local Similarity 66.7%; Pred. No. 82;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 YLSGADLNL 9
| || |||:
Db 107 YCSGGDLNV 115

RESULT 8
A49179
melanoma antigen homolog rpe1 - bovine (fragment)
C;Species: Bos primigenius taurus (cattle)

```
C;Date: 19-Dec-1993 #sequence_revision 18-Nov-1994 #text_change 09-Jul-2004
C;Accession: A49179; I45861
R;Kim, R.Y.; Wistow, G.J.
Exp. Eye Res. 55, 657-662, 1992
A;Title: The cDNA RPE1 and monoclonal antibody HMB-50 define gene products preferentiall
A;Reference number: A49179; MUID:93122163; PMID:1478275
A;Accession: A49179
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-491 <KIM>
A;Cross-references: UNIPROT:Q06154
A;Experimental source: retinal pigment epithelium
A;Note: sequence extracted from NCBI backbone (NCBIN:122438, NCBIP:122439)
C;Genetics:
A;Gene: RPE1

  Query Match             73.3%;  Score 33;  DB 2;  Length 491;
  Best Local Similarity   75.0%;  Pred. No. 1e+02;
  Matches    6;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 YLSGADLN 8
              ||:||||:
Db         93 YLAGADLS 100


RESULT 9
T23685
hypothetical protein M03B6.2 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C;Accession: T23685
R;Matthews, L.
submitted to the EMBL Data Library, August 1996
A;Reference number: Z19782
A;Accession: T23685
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-596 <WIL>
A;Cross-references: UNIPROT:Q93896; EMBL:Z78545; PIDN:CAB01766.1; GSPDB:GN00028; CESP:M0
A;Experimental source: clone M03B6
C;Genetics:
A;Gene: CESP:M03B6.2
A;Map position: X
A;Introns: 36/1; 94/1; 273/3; 320/1; 414/3; 533/1

  Query Match             73.3%;  Score 33;  DB 2;  Length 596;
  Best Local Similarity   87.5%;  Pred. No. 1.2e+02;
  Matches    7;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 YLSGADLN 8
              ||| ||||
Db        344 YLSKADLN 351


RESULT 10
T10585
serine proteinase homolog F9F13.80 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 09-Jul-2004
C;Accession: T10585
R;Bevan, M.; Pohl, T.; Weizenegger, T.; Bancroft, I.; Mewes, H.W.; Mayer, K.F.X.; Lemcke
submitted to the Protein Sequence Database, June 1999
A;Reference number: Z16991
A;Accession: T10585
A;Molecule type: DNA
A;Residues: 1-856 <BEV>
A;Cross-references: UNIPROT:Q9SUN6; EMBL:AL080253; GSPDB:GN00062; ATSP:F9F13.80
A;Experimental source: cultivar Columbia; BAC clone F9F13
C;Genetics:
A;Gene: ATSP:F9F13.80
A;Map position: 4
A;Introns: 84/2; 137/3; 271/2; 303/2; 327/3; 422/1; 533/3; 624/1; 718/1
C;Superfamily: subtilisin-like proteinase ag12; subtilisin homology

  Query Match             73.3%;  Score 33;  DB 2;  Length 856;
  Best Local Similarity   75.0%;  Pred. No. 1.8e+02;
  Matches    6;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 LSGADLNL 9
              :||:||||
Db        751 ISGSDLNL 758


RESULT 11
AI1604
weakly B. subtilis comG operon protein 7 (comGG) homolog lin1378 [imported] - Listeria in
C;Species: Listeria innocua
C;Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 09-Jul-2004
C;Accession: AI1604
R;Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloecker,
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.;
D.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A;Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Mat
ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,
A;Title: Comparative genomics of Listeria species.
A;Reference number: AB1077; MUID:21537279; PMID:11679669
A;Accession: AI1604
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-105 <GLA>
A;Cross-references: UNIPROT:Q92C13; GB:AL592022; PIDN:CAC96609.1; PID:g16413851; GSPDB:GN
A;Experimental source: strain Clip11262
C;Genetics:
A;Gene: lin1378

  Query Match             71.1%;  Score 32;  DB 2;  Length 105;
  Best Local Similarity   66.7%;  Pred. No. 31;
  Matches    6;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 YLSGADLNL 9
              ||: |:|||
Db         43 YLASAELNL 51


RESULT 12
S37313
transcription activator HlyU VC0678 [similarity] - Vibrio cholerae (strain N16961 serogrc
C;Species: Vibrio cholerae
C;Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 09-Jul-2004
C;Accession: S37313; H82292
R;Williams, S.G.; Attridge, S.R.; Manning, P.A.
Mol. Microbiol. 9, 751-760, 1993
A;Title: The transcriptional activator HlyU of Vibrio cholerae: nucleotide sequence and r
A;Reference number: S37312; MUID:94049116; PMID:8231807
A;Accession: S37313
A;Molecule type: DNA
A;Residues: 1-108 <WIL>
A;Cross-references: UNIPROT:P52695; EMBL:X66866; NID:g403330; PIDN:CAA47336.1; PID:g40333
R;Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.;
chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers, P.
l, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 406, 477-483, 2000
A;Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A;Reference number: A82035; MUID:20406833; PMID:10952301
A;Accession: H82292
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-108 <HEI>
A;Cross-references: GB:AE004154; GB:AE003852; NID:g9655115; PIDN:AAF93843.1; GSPDB:GN001
A;Experimental source: serogroup O1; strain N16961; biotype El Tor
C;Genetics:
A;Gene: hlyU; VC0678
A;Map position: 1
C;Superfamily: arsenical resistance operon repressor
```

```
  Query Match             71.1%;  Score 32;  DB 2;  Length 108;
  Best Local Similarity   66.7%;  Pred. No. 32;
  Matches    6;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 YLSGADLNL 9
              || || :||
Db          3 YLKGAPMNL 11


RESULT 13
A30169
hypothetical 19.9K protein (lktA 5' region) - Pasteurella haemolytica
C;Species: Pasteurella haemolytica
C;Date: 12-Oct-1989 #sequence_revision 12-Oct-1989 #text_change 09-Jul-2004
C;Accession: A30169; C35254
R;Highlander, S.K.; Chidambaram, M.; Engler, M.J.; Weinstock, G.M.
DNA 8, 15-28, 1989
A;Title: DNA sequence of the Pasteurella haemolytica leukotoxin gene cluster.
A;Reference number: A30169; MUID:89210283; PMID:2707120
A;Accession: A30169
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: DNA
A;Residues: 1-167 <HIG>
A;Cross-references: UNIPROT:P16533; GB:M24197; GB:M34943; GB:M34944; NID:g150511; PIDN:A
R;Highlander, S.K.; Engler, M.J.; Weinstock, G.M.
J. Bacteriol. 172, 2343-2350, 1990
A;Title: Secretion and expression of the Pasteurella haemolytica leukotoxin.
A;Reference number: A35254; MUID:90236888; PMID:2185213
A;Accession: C35254
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-20 <HI2>
A;Cross-references: GB:M24197; GB:M34943; GB:M34944
C;Superfamily: hemolysin C

  Query Match             71.1%;  Score 32;  DB 2;  Length 167;
  Best Local Similarity   77.8%;  Pred. No. 51;
  Matches    7;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 YLSGADLNL 9
              | | |||||
Db         55 YCSWADLNL 63


RESULT 14
S29515
lktC protein - Pasteurella haemolytica
C;Species: Pasteurella haemolytica
C;Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 09-Jul-2004
C;Accession: S29515
R;Lo, R.Y.C.; Strathdee, C.A.; Shewen, P.E.
Infect. Immun. 55, 1987-1996, 1987
A;Title: Nucleotide sequence of the leukotoxin genes of Pasteurella haemolytica A1.
A;Reference number: S29515; MUID:87306837; PMID:3040588
A;Accession: S29515
A;Molecule type: DNA
A;Residues: 1-167 <LOR>
A;Cross-references: UNIPROT:P16533; EMBL:M20730; NID:g150492; PIDN:AAA25528.1; PID:g1504
C;Genetics:
A;Gene: lktC
C;Superfamily: hemolysin C

  Query Match             71.1%;  Score 32;  DB 2;  Length 167;
  Best Local Similarity   77.8%;  Pred. No. 51;
  Matches    7;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 YLSGADLNL 9
              | | |||||
Db         55 YCSWADLNL 63


RESULT 15
F71064
micrococcal nuclease (EC 3.1.31.1) PH1212 precursor - Pyrococcus horikoshii
N;Alternate names: thermonuclease homolog
C;Species: Pyrococcus horikoshii
C;Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 12-Jul-2004
C;Accession: F71064
R;Kawarabayasi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; Sekine
 M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguchi,
DNA Res. 5, 55-76, 1998
A;Title: Complete sequence and gene organization of the genome of a hyper-thermophilic ar
A;Reference number: A71000; MUID:98344137; PMID:9679194
A;Accession: F71064
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-176 <KAW>
A;Cross-references: UNIPROT:O58971; GB:AP000005; NID:g3236132; PIDN:BAA30312.1; PID:g325
A;Experimental source: strain OT3
A;Note: this accession replaces an interim accession for a sequence replaced by GenBank
C;Genetics:
A;Gene: PH1212
C;Keywords: hydrolase
F;1-27/Domain: signal sequence #status predicted <SIG>

  Query Match             71.1%;  Score 32;  DB 2;  Length 176;
  Best Local Similarity   62.5%;  Pred. No. 53;
  Matches    5;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 YLSGADLN 8
              ||:| |:|
Db        128 YLNGTDIN 135


Search completed: May 17, 2005, 06:20:00
Job time : 14.25 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: May 17, 2005, 06:13:14 ; Search time 51.75 Seconds
(without alignments)
89.057 Million cell updates/sec

Title: US-10-725-373-2
Perfect score: 45
Sequence: 1 YLSGADLNL 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : UniProt_03:*
1: uniprot_sprot:*
2: uniprot_trembl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	40	88.9	420	2	Q68DM9	Q68dm9 homo sapien
2	40	88.9	702	1	CEA5_HUMAN	P06731 homo sapien
3	40	88.9	702	2	Q8N4D0	Q8n4d0 homo sapien
4	37	82.2	499	2	Q6PW77	Q6pw77 acremonium
5	37	82.2	856	2	Q7NMP7	Q7nmp7 gloeobacter
6	36	80.0	237	2	Q7U4R6	Q7u4r6 synechococc
7	36	80.0	284	2	Q7NL68	Q7nl68 gloeobacter
8	36	80.0	322	2	Q8A067	Q8a067 bacteroides
9	36	80.0	828	2	Q87DG0	Q87dg0 xylella fas
10	36	80.0	834	2	Q9PD71	Q9pd71 xylella fas
11	36	80.0	953	2	Q8YX84	Q8yx84 anabaena sp
12	35	77.8	186	2	Q6CHW6	Q6chw6 yarrowia li
13	35	77.8	316	2	Q9CF54	Q9cf54 lactococcus
14	35	77.8	519	2	Q47916	Q47916 fibrobacter
15	35	77.8	818	2	Q6C363	Q6c363 yarrowia li
16	35	77.8	1770	1	N4B2_HUMAN	Q86uw6 homo sapien
17	34	75.6	144	2	Q7P798	Q7p798 fusobacteri
18	34	75.6	154	2	Q7V2V0	Q7v2v0 prochloroco
19	34	75.6	340	2	Q72VL5	Q72vl5 leptospira
20	34	75.6	340	2	Q8F971	Q8f971 leptospira
21	34	75.6	348	2	Q8NSD4	Q8nsd4 corynebacte
22	34	75.6	362	2	Q6M738	Q6m738 corynebacte
23	34	75.6	455	1	PEX3_PICPA	Q92262 pichia past
24	34	75.6	457	2	Q7NCT4	Q7nct4 gloeobacter
25	34	75.6	480	2	Q7YT18	Q7yt18 strongyloce
26	34	75.6	480	2	Q9GUY6	Q9guy6 hemicentrot
27	34	75.6	496	2	Q9BZV2	Q9bzv2 homo sapien
28	34	75.6	506	2	Q9PKP0	Q9pkp0 chlamydia m
29	34	75.6	540	2	O31716	O31716 bacillus su
30	34	75.6	560	2	Q8TPL6	Q8tpl6 methanosarc
31	34	75.6	686	2	Q6LYF2	Q6lyf2 methanococc
32	34	75.6	763	2	O62831	O62831 bos taurus
33	34	75.6	773	2	O62828	O62828 bos taurus
34	34	75.6	935	2	Q6N7G0	Q6n7g0 rhodopseudo
35	34	75.6	938	2	Q965V4	Q965v4 caenorhabdi
36	33	73.3	131	2	Q7TV15	Q7tv15 prochloroco
37	33	73.3	176	2	Q6QJ36	Q6qj36 arabidopsis
38	33	73.3	235	2	Q87FL4	Q87fl4 vibrio para
39	33	73.3	248	2	Q6ND53	Q6nd53 rhodopseudo
40	33	73.3	248	2	Q89WW2	Q89ww2 bradyrhizob
41	33	73.3	286	2	Q7YL37	Q7yl37 schmardaea
42	33	73.3	333	2	Q7QER9	Q7qer9 anopheles g
43	33	73.3	349	2	Q8ZWB6	Q8zwb6 pyrobaculum
44	33	73.3	352	2	Q65258	Q65258 african swi
45	33	73.3	388	2	Q948C5	Q948c5 oryza sativ

ALIGNMENTS

RESULT 1
Q68DM9
ID Q68DM9 PRELIMINARY; PRT; 420 AA.
AC Q68DM9;
DT 25-OCT-2004 (TrEMBLrel. 28, Created)
DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE Hypothetical protein DKFZp781M2392.
GN Name=DKFZp781M2392;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Colon carcinoma;
RG The German cDNA Consortium;
RA Poustka A., Albert R., Moosmayer P., Schupp I., Wellenreuther R.,
RA Mewes H.W., Weil B., Amid C., Osanger A., Fobo G., Han M., Wiemann S.;
RL Submitted (AUG-2004) to the EMBL/GenBank/DDBJ databases.
DR EMBL; CR749337; CAH18191.1; -.
DR InterPro; IPR001589; Actbind_actnin.
DR InterPro; IPR003599; Ig.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003598; Ig_c2.
DR Pfam; PF00047; ig; 3.
DR SMART; SM00409; IG; 3.
DR SMART; SM00408; IGc2; 3.
DR PROSITE; PS00019; ACTININ_1; UNKNOWN_1.
DR PROSITE; PS50835; IG_LIKE; 3.
KW Hypothetical protein.
SQ SEQUENCE 420 AA; 45508 MW; 6E30C0B4A00D0F59 CRC64;

Query Match 88.9%; Score 40; DB 2; Length 420;
Best Local Similarity 88.9%; Pred. No. 15;
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 YLSGADLNL 9
              |||||:|||
Db        323 YLSGANLNL 331
```

RESULT 2
CEA5_HUMAN
ID CEA5_HUMAN STANDARD; PRT; 702 AA.
AC P06731;
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 25-OCT-2004 (Rel. 45, Last annotation update)
DE Carcinoembryonic antigen-related cell adhesion molecule 5 precursor
DE (Carcinoembryonic antigen) (CEA) (Meconium antigen 100) (CD66e
DE antigen).
GN Name=CEACAM5; Synonyms=CEA;
OS Homo sapiens (Human).

```
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=90258861; PubMed=2342461;
RA   Schrewe H., Thompson J., Bona M., Hefta L.J.F., Maruya A.,
RA   Hassauer M., Shively J.E., von Kleist S., Zimmermann W.;
RT   "Cloning of the complete gene for carcinoembryonic antigen: analysis
RT   of its promoter indicates a region conveying cell type-specific
RT   expression.";
RL   Mol. Cell. Biol. 10:2738-2748(1990).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=88038876; PubMed=3670312;
RA   Beauchemin N., Benchimol S., Cournoyer D., Fuks A., Stanners C.P.;
RT   "Isolation and characterization of full-length functional cDNA clones
RT   for human carcinoembryonic antigen.";
RL   Mol. Cell. Biol. 7:3221-3230(1987).
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=89122014; PubMed=3220478;
RA   Barnett T., Goebel S.J., Nothdurft M.A., Elting J.J.;
RT   "Carcinoembryonic antigen family: characterization of cDNAs coding for
RT   NCA and CEA and suggestion of nonrandom sequence variation in their
RT   conserved loop-domains.";
RL   Genomics 3:59-66(1988).
RN   [4]
RP   SEQUENCE OF 5-702 FROM N.A.
RX   MEDLINE=87128144; PubMed=3814146;
RA   Oikawa S., Nakazato H., Kosaki G.;
RT   "Primary structure of human carcinoembryonic antigen (CEA) deduced
RT   from cDNA sequence.";
RL   Biochem. Biophys. Res. Commun. 142:511-518(1987).
RN   [5]
RP   SEQUENCE OF 331-702 FROM N.A.
RX   MEDLINE=87204247; PubMed=3033671;
RA   Zimmermann W., Ortlieb B., Friedrich R., von Kleist S.;
RT   "Isolation and characterization of cDNA clones encoding the human
RT   carcinoembryonic antigen reveal a highly conserved repeating
RT   structure.";
RL   Proc. Natl. Acad. Sci. U.S.A. 84:2960-2964(1987).
CC   -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.
CC   -!- TISSUE SPECIFICITY: Found in adenocarcinomas of endodermally
CC       derived digestive system epithelium and fetal colon.
CC   -!- PTM: Complex immunoreactive glycoprotein with a MW of 180 kDa
CC       comprising 60% carbohydrate.
CC   -!- SIMILARITY: Belongs to the immunoglobulin superfamily. CEA family.
CC   -!- SIMILARITY: Contains 7 immunoglobulin-like domains.
CC   -!- DATABASE: NAME=PROW; NOTE=CD guide CD66e entry;
CC       WWW="http://www.ncbi.nlm.nih.gov/prow/cd/cd66e.htm".
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M17303; AAB59513.1; -.
DR   EMBL; M59262; AAA62835.1; ALT_SEQ.
DR   EMBL; M59255; AAA62835.1; JOINED.
DR   EMBL; M59256; AAA62835.1; JOINED.
DR   EMBL; M59257; AAA62835.1; JOINED.
DR   EMBL; M59258; AAA62835.1; JOINED.
DR   EMBL; M59259; AAA62835.1; JOINED.
DR   EMBL; M59260; AAA62835.1; JOINED.
DR   EMBL; M59261; AAA62835.1; JOINED.
DR   EMBL; M59709; -; NOT_ANNOTATED_CDS.
DR   EMBL; M59710; -; NOT_ANNOTATED_CDS.
DR   EMBL; M29540; AAA51967.1; -.
DR   EMBL; X16455; CAA34474.1; -.
DR   EMBL; M15042; AAA51963.1; -.
DR   EMBL; M16234; AAA51972.1; -.
DR   PIR; A36319; A36319.
DR   PDB; 1E07; Model; A=35-676.
DR   Genew; HGNC:1817; CEACAM5.
DR   MIM; 114890; -.
DR   GO; GO:0005887; C:integral to plasma membrane; TAS.
DR   InterPro; IPR007110; Ig-like.
DR   Pfam; PF00047; ig; 6.
DR   PROSITE; PS50835; IG_LIKE; 6.
KW   3D-structure; Glycoprotein; GPI-anchor; Immunoglobulin domain;
KW   Lipoprotein; Membrane; Repeat; Signal.
FT   SIGNAL        1     34
FT   CHAIN        35    685       Carcinoembryonic antigen-related cell
FT                                 adhesion molecule 5.
FT   PROPEP      686    702       Removed in mature form (Potential).
FT   DOMAIN       35    144       Ig-like 1.
FT   DOMAIN      146    237       Ig-like 2.
FT   DOMAIN      238    322       Ig-like 3.
FT   DOMAIN      324    415       Ig-like 4.
FT   DOMAIN      416    498       Ig-like 5.
FT   DOMAIN      502    593       Ig-like 6.
FT   DOMAIN      594    677       Ig-like 7.
FT   CARBOHYD    104    104       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    115    115       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    152    152       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    182    182       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    197    197       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    204    204       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    208    208       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    246    246       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    256    256       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    274    274       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    288    288       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    292    292       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    309    309       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    330    330       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    351    351       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    360    360       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    375    375       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    432    432       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    466    466       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    480    480       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    508    508       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    529    529       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    553    553       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    560    560       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    580    580       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    612    612       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    650    650       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    665    665       N-linked (GlcNAc. . .) (Potential).
FT   LIPID       685    685       GPI-anchor amidated alanine (Potential).
FT   CONFLICT    320    320       Missing (in Ref. 4).
SQ   SEQUENCE   702 AA;  76795 MW;  6299AE26CDDBDB5C CRC64;

  Query Match             88.9%;  Score 40;  DB 1;  Length 702;
  Best Local Similarity   88.9%;  Pred. No. 25;
  Matches    8;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 YLSGADLNL 9
              |||||:|||
Db        605 YLSGANLNL 613


RESULT 3
Q8N4D0
ID   Q8N4D0      PRELIMINARY;      PRT;   702 AA.
AC   Q8N4D0;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   CEACAM5 protein.
OS   Homo sapiens (Human).
```

```
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Colon;
RX   MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA   Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
RA   Jones S.J., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length human
RT   and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Colon;
RA   Strausberg R.;
RL   Submitted (JUL-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC034671; AAH34671.1; -.
DR   HSSP; Q61353; 1L6Z.
DR   InterPro; IPR001589; Actbind_actnin.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR003598; Ig_c2.
DR   Pfam; PF00047; ig; 6.
DR   SMART; SM00408; IGc2; 3.
DR   PROSITE; PS00019; ACTININ_1; UNKNOWN_3.
DR   PROSITE; PS50835; IG_LIKE; 6.
SQ   SEQUENCE   702 AA;  76781 MW;  97CCFB7399A0B05A CRC64;

  Query Match             88.9%;  Score 40;  DB 2;  Length 702;
  Best Local Similarity   88.9%;  Pred. No. 25;
  Matches    8;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 YLSGADLNL 9
              |||||:|||
Db        605 YLSGANLNL 613


RESULT 4
Q6PW77
ID   Q6PW77      PRELIMINARY;      PRT;   499 AA.
AC   Q6PW77;
DT   05-JUL-2004 (TrEMBLrel. 27, Created)
DT   05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT   05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE   Glucooligosaccharide oxidase.
OS   Acremonium strictum (Black bundle disease fungus).
OC   Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC   Hypocreomycetidae; Hypocreales; Hypocreaceae; mitosporic Hypocreaceae;
OC   Acremonium.
OX   NCBI_TaxID=5046;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Lee M.-H., Lai W.-L., Lin S.-F., Liaw S.-H., Tsai Y.-C.;
RL   Submitted (MAR-2004) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY573966; AAS79317.1; -.
DR   GO; GO:0006118; P:electron transport; IEA.
DR   InterPro; IPR006094; Oxid_FAD_bind_N.
DR   InterPro; IPR006093; Oxred_FAD_BS.
DR   Pfam; PF01565; FAD_binding_4; 1.
DR   PROSITE; PS00862; OX2_COVAL_FAD; UNKNOWN_1.
SQ   SEQUENCE   499 AA;  55237 MW;  BCB56CF1F4E922CE CRC64;

  Query Match             82.2%;  Score 37;  DB 2;  Length 499;
  Best Local Similarity   77.8%;  Pred. No. 74;
  Matches    7;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 YLSGADLNL 9
              || |||||:
Db        323 YLYGADLNI 331


RESULT 5
Q7NMP7
ID   Q7NMP7      PRELIMINARY;      PRT;   856 AA.
AC   Q7NMP7;
DT   01-MAR-2004 (TrEMBLrel. 26, Created)
DT   01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Two-component hybrid sensor and regulator.
GN   OrderedLocusNames=glr0718;
OS   Gloeobacter violaceus.
OC   Bacteria; Cyanobacteria; Chroococcales; Gloeobacter.
OX   NCBI_TaxID=33072;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=PCC 7421;
RX   MEDLINE=22977040; PubMed=14621292;
RA   Nakamura Y., Kaneko T., Sato S., Mimuro M., Miyashita H., Tsuchiya T.,
RA   Sasamoto S., Watanabe A., Kawashima K., Kishida Y., Kiyokawa C.,
RA   Kohara M., Matsumoto M., Matsuno A., Nakazaki N., Shimpo S.,
RA   Takeuchi C., Yamada M., Tabata S.;
RT   "Complete genome structure of Gloeobacter violaceus PCC 7421, a
RT   cyanobacterium that lacks thylakoids.";
RL   DNA Res. 10:137-145(2003).
CC   -!- SIMILARITY: Contains 1 histidine kinase domain.
DR   EMBL; AP006570; BAC88659.1; -.
DR   HSSP; P39928; 1OXK.
DR   GO; GO:0016020; C:membrane; IEA.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0003677; F:DNA binding; IEA.
DR   GO; GO:0016301; F:kinase activity; IEA.
DR   GO; GO:0016740; F:transferase activity; IEA.
DR   GO; GO:0000156; F:two-component response regulator activity; IEA.
DR   GO; GO:0000155; F:two-component sensor molecule activity; IEA.
DR   GO; GO:0007600; P:sensory perception; IEA.
DR   GO; GO:0000160; P:two-component signal transduction system (p. . .; IEA.
DR   InterPro; IPR003594; ATPbind_ATPase.
DR   InterPro; IPR004358; Bact_sens_pr_C.
DR   InterPro; IPR011006; CheY_like.
DR   InterPro; IPR003018; GAF.
DR   InterPro; IPR005467; His_kinase.
DR   InterPro; IPR003661; His_kinA_N.
DR   InterPro; IPR009082; His_kin_homodim.
DR   InterPro; IPR001610; PAC.
DR   InterPro; IPR000014; PAS.
DR   InterPro; IPR000700; PAS-assoc_C.
DR   InterPro; IPR001789; Response_reg.
DR   Pfam; PF01590; GAF; 1.
DR   Pfam; PF02518; HATPase_c; 1.
DR   Pfam; PF00512; HisKA; 1.
DR   Pfam; PF00785; PAC; 1.
DR   Pfam; PF00072; Response_reg; 1.
DR   PRINTS; PR00344; BCTRLSENSOR.
DR   ProDom; PD000039; Response_reg; 1.
DR   TIGRFAMs; TIGR00229; sensory_box; 1.
DR   PROSITE; PS50109; HIS_KIN; 1.
DR   PROSITE; PS50113; PAC; 1.
DR   PROSITE; PS50110; RESPONSE_REGULATORY; 1.
KW   Complete proteome; Kinase; Phosphorylation; Sensory transduction;
KW   Transferase.
SQ   SEQUENCE   856 AA;  94436 MW;  1D0C91229B3CAEF1 CRC64;
```

```
  Query Match             82.2%;  Score 37;  DB 2;  Length 856;
  Best Local Similarity   77.8%;  Pred. No. 1.3e+02;
  Matches    7;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 YLSGADLNL 9
              ||:||||:|
Db         98 YLNGADLHL 106


RESULT 6
Q7U4R6
ID   Q7U4R6      PRELIMINARY;      PRT;   237 AA.
AC   Q7U4R6;
DT   01-OCT-2003 (TrEMBLrel. 25, Created)
DT   01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Hypothetical protein.
GN   OrderedLocusNames=SYNW1998;
OS   Synechococcus sp. (strain WH8102).
OC   Bacteria; Cyanobacteria; Chroococcales; Synechococcus.
OX   NCBI_TaxID=84588;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=22825697; PubMed=12917641; DOI=10.1038/nature01943;
RA   Palenik B., Brahamsha B., Larimer F.W., Land M.L., Hauser L.,
RA   Chain P., Lamerdin J.E., Regala W., Allen E.E., McCarren J.,
RA   Paulsen I.T., Dufresne A., Partensky F., Webb E.A., Waterbury J.;
RT   "The genome of a motile marine Synechococcus.";
RL   Nature 424:1037-1042(2003).
DR   EMBL; BX569694; CAE08513.1; -.
DR   InterPro; IPR001646; 5peptide_repeat.
DR   Pfam; PF00805; Pentapeptide; 3.
KW   Complete proteome; Hypothetical protein.
SQ   SEQUENCE   237 AA;  24963 MW;  9D7EF9A986A434FB CRC64;

  Query Match             80.0%;  Score 36;  DB 2;  Length 237;
  Best Local Similarity   87.5%;  Pred. No. 56;
  Matches    7;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 YLSGADLN 8
              |||||||:
Db        125 YLSGADLS 132


RESULT 7
Q7NL68
ID   Q7NL68      PRELIMINARY;      PRT;   284 AA.
AC   Q7NL68;
DT   01-MAR-2004 (TrEMBLrel. 26, Created)
DT   01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Gll1258 protein.
GN   OrderedLocusNames=gll1258;
OS   Gloeobacter violaceus.
OC   Bacteria; Cyanobacteria; Chroococcales; Gloeobacter.
OX   NCBI_TaxID=33072;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=PCC 7421;
RX   MEDLINE=22977040; PubMed=14621292;
RA   Nakamura Y., Kaneko T., Sato S., Mimuro M., Miyashita H., Tsuchiya T.,
RA   Sasamoto S., Watanabe A., Kawashima K., Kishida Y., Kiyokawa C.,
RA   Kohara M., Matsumoto M., Matsuno A., Nakazaki N., Shimpo S.,
RA   Takeuchi C., Yamada M., Tabata S.;
RT   "Complete genome structure of Gloeobacter violaceus PCC 7421, a
RT   cyanobacterium that lacks thylakoids.";
RL   DNA Res. 10:137-145(2003).
DR   EMBL; AP006572; BAC89199.1; -.
DR   InterPro; IPR001646; 5peptide_repeat.
DR   Pfam; PF00805; Pentapeptide; 4.
KW   Complete proteome.
SQ   SEQUENCE   284 AA;  29498 MW;  F4485B9BB089E1BE CRC64;

  Query Match             80.0%;  Score 36;  DB 2;  Length 284;
  Best Local Similarity   87.5%;  Pred. No. 67;
  Matches    7;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 YLSGADLN 8
              |||||||:
Db        178 YLSGADLS 185


RESULT 8
Q8A067
ID   Q8A067      PRELIMINARY;      PRT;   322 AA.
AC   Q8A067;
DT   01-JUN-2003 (TrEMBLrel. 24, Created)
DT   01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Chitin deacetylase.
GN   OrderedLocusNames=BT4154;
OS   Bacteroides thetaiotaomicron.
OC   Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC   Bacteroidaceae; Bacteroides.
OX   NCBI_TaxID=818;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=VPI-5482 / ATCC 29148;
RX   MEDLINE=22550858; PubMed=12663928; DOI=10.1126/science.1080029;
RA   Xu J., Bjursell M.K., Himrod J., Deng S., Carmichael L.K.,
RA   Chiang H.C., Hooper L.V., Gordon J.I.;
RT   "A genomic view of the human-Bacteroides thetaiotaomicron symbiosis.";
RL   Science 299:2074-2076(2003).
DR   EMBL; AE016944; AAO79259.1; -.
DR   GO; GO:0016810; F:hydrolase activity, acting on carbon-nitrog. . .; IEA.
DR   GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR   InterPro; IPR002509; Polysac_deacet.
DR   Pfam; PF01522; Polysacc_deac_1; 1.
KW   Complete proteome.
SQ   SEQUENCE   322 AA;  37211 MW;  8A5A40FEFFBB1145 CRC64;

  Query Match             80.0%;  Score 36;  DB 2;  Length 322;
  Best Local Similarity   77.8%;  Pred. No. 76;
  Matches    7;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 YLSGADLNL 9
              |||||| |:
Db         17 YLSGADWNV 25


RESULT 9
Q87DG0
ID   Q87DG0      PRELIMINARY;      PRT;   828 AA.
AC   Q87DG0;
DT   01-JUN-2003 (TrEMBLrel. 24, Created)
DT   01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Hypothetical protein.
GN   OrderedLocusNames=PD0725;
OS   Xylella fastidiosa (strain Temecula1 / ATCC 700964).
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC   Xanthomonadaceae; Xylella.
OX   NCBI_TaxID=183190;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=22421331; PubMed=12533478;
RX   DOI=10.1128/JB.185.3.1018-1026.2003;
RA   Van Sluys M.A., de Oliveira M.C., Monteiro-Vitorello C.B.,
RA   Miyaki C.Y., Furlan L.R., Camargo L.E.A., da Silva A.C.R., Moon D.H.,
RA   Takita M.A., Lemos E.G.M., Machado M.A., Ferro M.I.T., da Silva F.R.,
RA   Goldman M.H.S., Goldman G.H., Lemos M.V.F., El-Dorry H., Tsai S.M.,
RA   Carrer H., Carraro D.M., de Oliveira R.C., Nunes L.R., Siqueira W.J.,
RA   Coutinho L.L., Kimura E.T., Ferro E.S., Harakava R., Kuramae E.E.,
RA   Marino C.L., Giglioti E., Abreu I.L., Alves L.M.C., do Amaral A.M.,
RA   Baia G.S., Blanco S.R., Brito M.S., Cannavan F.S., Celestino A.V.,
```

```
RA   da Cunha A.F., Fenille R.C., Ferro J.A., Formighieri E.F., Kishi L.T.,
RA   Leoni S.G., Oliveira A.R., Rosa V.E. Jr., Sassaki F.T., Sena J.A.D.,
RA   de Souza A.A., Truffi D., Tsukumo F., Yanai G.M., Zaros L.G.,
RA   Civerolo E.L., Simpson A.J.G., Almeida N.F. Jr., Setubal J.C.,
RA   Kitajima J.P.;
RT   "Comparative analyses of the complete genome sequences of Pierce's
RT   disease and citrus variegated chlorosis strains of Xylella
RT   fastidiosa.";
RL   J. Bacteriol. 185:1018-1026(2003).
DR   EMBL; AE012555; AAO28594.1; -.
KW   Complete proteome.
SQ   SEQUENCE   828 AA;  91728 MW;  CEEF57878F0E09FC CRC64;

  Query Match             80.0%;  Score 36;  DB 2;  Length 828;
  Best Local Similarity   87.5%;  Pred. No. 2e+02;
  Matches    7;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 YLSGADLN 8
              |||| |||
Db        388 YLSGMDLN 395


RESULT 10
Q9PD71
ID   Q9PD71      PRELIMINARY;      PRT;   834 AA.
AC   Q9PD71;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Hypothetical protein.
GN   OrderedLocusNames=Xf1508;
OS   Xylella fastidiosa.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC   Xanthomonadaceae; Xylella.
OX   NCBI_TaxID=2371;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=9a5c;
RX   MEDLINE=20365717; PubMed=10910347; DOI=10.1038/35018003;
RA   Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,
RA   Alvarenga R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,
RA   Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,
RA   Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carrer H.,
RA   Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,
RA   Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorry H.,
RA   Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,
RA   Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furlan L.R.,
RA   Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
RA   Ho P.L., Hoheisel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,
RA   Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,
RA   Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
RA   Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,
RA   Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,
RA   Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,
RA   Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
RA   Nhani A. Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,
RA   de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
RA   Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,
RA   Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
RA   de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,
RA   da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,
RA   da Silveira J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,
RA   de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tsuhako M.H.,
RA   Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
RA   Zago M.A., Zatz M., Meidanis J., Setubal J.C.;
RT   "The genome sequence of the plant pathogen Xylella fastidiosa.";
RL   Nature 406:151-159(2000).
DR   EMBL; AE003980; AAF84317.1; -.
DR   PIR; F82673; F82673.
KW   Complete proteome; Hypothetical protein.
SQ   SEQUENCE   834 AA;  92558 MW;  A424E3245CA07470 CRC64;

  Query Match             80.0%;  Score 36;  DB 2;  Length 834;
  Best Local Similarity   87.5%;  Pred. No. 2e+02;
  Matches    7;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 YLSGADLN 8
              |||| |||
Db        391 YLSGMDLN 398


RESULT 11
Q8YX84
ID   Q8YX84      PRELIMINARY;      PRT;   953 AA.
AC   Q8YX84;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Alr1331 protein.
GN   OrderedLocusNames=alr1331;
OS   Anabaena sp. (strain PCC 7120).
OC   Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
OX   NCBI_TaxID=103690;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21595285; PubMed=11759840;
RA   Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,
RA   Watanabe A., Iriguchi M., Ishikawa A., Kawashima K., Kimura T.,
RA   Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
RA   Nakazaki N., Shimpo S., Sugimoto M., Takazawa M., Yamada M.,
RA   Yasuda M., Tabata S.;
RT   "Complete genomic sequence of the filamentous nitrogen-fixing
RT   cyanobacterium Anabaena sp. strain PCC 7120.";
RL   DNA Res. 8:205-213(2001).
DR   EMBL; AP003585; BAB73288.1; -.
DR   PIR; AH1972; AH1972.
DR   Pfam; PF00805; Pentapeptide; 2.
KW   Complete proteome.
SQ   SEQUENCE   953 AA;  108028 MW;  AF08552CDE7DC724 CRC64;

  Query Match             80.0%;  Score 36;  DB 2;  Length 953;
  Best Local Similarity   87.5%;  Pred. No. 2.3e+02;
  Matches    7;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 YLSGADLN 8
              |||||||:
Db        828 YLSGADLS 835


RESULT 12
Q6CHW6
ID   Q6CHW6      PRELIMINARY;      PRT;   186 AA.
AC   Q6CHW6;
DT   25-OCT-2004 (TrEMBLrel. 28, Created)
DT   25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT   25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE   Similarity.
GN   ORFNames=YALI0A03993g;
OS   Yarrowia lipolytica CLIB99.
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Dipodascaceae; Yarrowia.
OX   NCBI_TaxID=284591;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CLIB99;
RG   Genolevures;
RA   Dujon B., Sherman D., Fischer G., Durrens P., Casaregola S.,
RA   Lafontaine I., de Montigny J., Marck C., Neuveglise C., Talla E.,
RA   Goffard N., Frangeul L., Aigle M., Anthouard V., Babour A., Barbe V.,
RA   Barnay S., Blanchin S., Beckerich J.M., Beyne E., Bleykasten C.,
RA   Boisrame A., Boyer J., Cattolico L., Confanioleri F., de Daruvar A.,
RA   Despons L., Fabre E., Fairhead C., Ferry-Dumazet H., Groppi A.,
RA   Hantraye F., Hennequin C., Jauniaux N., Joyet P., Kachouri R.,
RA   Kerrest A., Koszul R., Lemaire M., Lesur I., Ma L., Muller H.,
RA   Nicaud J.M., Nikolski M., Oztas S., Ozier-Kalogeropoulos O.,
RA   Pellenz S., Potier S., Richard G.F., Straub M.L., Suleau A.,
```

```
RA   Swennene D., Tekaia F., Wesolowski-Louvel M., Westhof E., Wirth B.,
RA   Zeniou-Meyer M., Zivanovic I., Bolotin-Fukuhara M., Thierry A.,
RA   Bouchier C., Caudron B., Scarpelli C., Gaillardin C., Weissenbach J.,
RA   Wincker P., Souciet J.L.;
RT   "Genome evolution in yeasts.";
RL   Nature 430:35-44(2004).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CLIB99;
RA   Genoscope;
RL   Submitted (JUL-2004) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; CR382127; CAG83668.1; -.
DR   GO; GO:0005515; F:protein binding; IEA.
DR   InterPro; IPR000210; BTB_POZ.
DR   PROSITE; PS50097; BTB; 1.
SQ   SEQUENCE   186 AA;  20853 MW;  BC0AA9EE2C291511 CRC64;

  Query Match             77.8%;  Score 35;  DB 2;  Length 186;
  Best Local Similarity   77.8%;  Pred. No. 71;
  Matches    7;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 YLSGADLNL 9
              || | ||||
Db         79 YLHGQDLNL 87


RESULT 13
Q9CF54
ID   Q9CF54      PRELIMINARY;      PRT;   316 AA.
AC   Q9CF54;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   2-dehydro-3-deoxygluconokinase (EC 2.7.1.45).
GN   Name=kdgK; OrderedLocusNames=LL1627;
OS   Lactococcus lactis (subsp. lactis) (Streptococcus lactis).
OC   Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Lactococcus.
OX   NCBI_TaxID=1360;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=IL1403;
RX   MEDLINE=21235186; PubMed=11337471; DOI=10.1101/gr.GR-1697R;
RA   Bolotin A., Wincker P., Mauger S., Jaillon O., Malarme K.,
RA   Weissenbach J., Ehrlich S.D., Sorokin A.;
RT   "The complete genome sequence of the lactic acid bacterium Lactococcus
RT   lactis ssp. lactis IL1403.";
RL   Genome Res. 11:731-753(2001).
DR   EMBL; AE006393; AAK05725.1; -.
DR   PIR; C86828; C86828.
DR   HSSP; P05054; 1RKD.
DR   GO; GO:0008673; F:2-dehydro-3-deoxygluconokinase activity; IEA.
DR   GO; GO:0016301; F:kinase activity; IEA.
KW   Complete proteome; Kinase.
SQ   SEQUENCE   316 AA;  34375 MW;  698732D1BD796CB1 CRC64;

  Query Match             77.8%;  Score 35;  DB 2;  Length 316;
  Best Local Similarity   66.7%;  Pred. No. 1.2e+02;
  Matches    6;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 YLSGADLNL 9
              ||:||:||:
Db         32 YLAGAELNV 40


RESULT 14
Q47916
ID   Q47916      PRELIMINARY;      PRT;   519 AA.
AC   Q47916;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Endoglucanase CelG.
GN   Name=celG;
OS   Fibrobacter succinogenes (Bacteroides succinogenes).
OC   Bacteria; Fibrobacteres; Fibrobacterales; Fibrobacteraceae;
OC   Fibrobacter.
OX   NCBI_TaxID=833;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S85;
RX   MEDLINE=97017599; PubMed=8864216;
RA   Iyo A.H., Forsberg C.W.;
RT   "Endoglucanase G from Fibrobacter succinogenes S85 belongs to a class
RT   of enzymes characterized by a basic C-terminal domain.";
RL   Can. J. Microbiol. 42:934-943(1996).
DR   EMBL; U33887; AAB38548.1; -.
DR   HSSP; P17901; 1EDG.
DR   GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . .; IEA.
DR   GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR   InterPro; IPR001547; Glyco_hydro_5.
DR   Pfam; PF00150; Cellulase; 1.
SQ   SEQUENCE   519 AA;  56848 MW;  B06D2113B10FF27E CRC64;

  Query Match             77.8%;  Score 35;  DB 2;  Length 519;
  Best Local Similarity   87.5%;  Pred. No. 2e+02;
  Matches    7;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 LSGADLNL 9
              |||:||||
Db        334 LSGSDLNL 341


RESULT 15
Q6C363
ID   Q6C363      PRELIMINARY;      PRT;   818 AA.
AC   Q6C363;
DT   25-OCT-2004 (TrEMBLrel. 28, Created)
DT   25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT   25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE   Similar to ca|CaNAM2 Candida albicans mitochondrial leucinetRNA
DE   ligase.
GN   ORFNames=YALI0F02299g;
OS   Yarrowia lipolytica CLIB99.
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Dipodascaceae; Yarrowia.
OX   NCBI_TaxID=284591;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CLIB99;
RG   Genolevures;
RA   Dujon B., Sherman D., Fischer G., Durrens P., Casaregola S.,
RA   Lafontaine I., de Montigny J., Marck C., Neuveglise C., Talla E.,
RA   Goffard N., Frangeul L., Aigle M., Anthouard V., Babour A., Barbe V.,
RA   Barnay S., Blanchin S., Beckerich J.M., Beyne E., Bleykasten C.,
RA   Boisrame A., Boyer J., Cattolico L., Confanioleri F., de Daruvar A.,
RA   Despons L., Fabre E., Fairhead C., Ferry-Dumazet H., Groppi A.,
RA   Hantraye F., Hennequin C., Jauniaux N., Joyet P., Kachouri R.,
RA   Kerrest A., Koszul R., Lemaire M., Lesur I., Ma L., Muller H.,
RA   Nicaud J.M., Nikolski M., Oztas S., Ozier-Kalogeropoulos O.,
RA   Pellenz S., Potier S., Richard G.F., Straub M.L., Suleau A.,
RA   Swennene D., Tekaia F., Wesolowski-Louvel M., Westhof E., Wirth B.,
RA   Zeniou-Meyer M., Zivanovic I., Bolotin-Fukuhara M., Thierry A.,
RA   Bouchier C., Caudron B., Scarpelli C., Gaillardin C., Weissenbach J.,
RA   Wincker P., Souciet J.L.;
RT   "Genome evolution in yeasts.";
RL   Nature 430:35-44(2004).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CLIB99;
RA   Genoscope;
RL   Submitted (JUL-2004) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; CR382132; CAG77701.1; -.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0004823; F:leucine-tRNA ligase activity; IEA.
DR   GO; GO:0016874; F:ligase activity; IEA.
DR   GO; GO:0006429; P:leucyl-tRNA aminoacylation; IEA.
```

```
DR   InterPro; IPR002302; Leu-tRNAsynt1a.
DR   InterPro; IPR002300; tRNA-synt_1a.
DR   InterPro; IPR001412; tRNA-synt_I.
DR   InterPro; IPR009080; tRNAsyn_1a_bind.
DR   Pfam; PF00133; tRNA-synt_1; 1.
DR   PRINTS; PR00985; TRNASYNTHLEU.
DR   TIGRFAMs; TIGR00396; leuS_bact; 1.
DR   PROSITE; PS00178; AA_TRNA_LIGASE_I; 1.
KW   Ligase.
SQ   SEQUENCE   818 AA;  91359 MW;  3AC35304306490AE CRC64;

  Query Match             77.8%;  Score 35;  DB 2;  Length 818;
  Best Local Similarity   87.5%;  Pred. No. 3.1e+02;
  Matches    7;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 YLSGADLN 8
              | ||||||
Db        514 YWSGADLN 521


Search completed: May 17, 2005, 06:23:34
Job time : 55.75 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: May 17, 2005, 06:13:14 ; Search time 66 Seconds
(without alignments)
52.740 Million cell updates/sec

Title: US-10-725-373-2
Perfect score: 45
Sequence: 1 YLSGADLNL 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_16Dec04:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	45	100.0	9	2	AAY09526	Aay09526 Carcinoem
2	45	100.0	9	3	AAB13750	Aab13750 Peptide f
3	45	100.0	9	4	AAB97818	Aab97818 Carcinoem
4	45	100.0	9	4	AAE05124	Aae05124 Modified
5	45	100.0	9	5	AAB47917	Aab47917 Modified
6	45	100.0	9	5	AAE19089	Aae19089 HLA-A24 r
7	45	100.0	9	8	ADN63713	Adn63713 HLA bindi
8	45	100.0	701	4	AAB97817	Aab97817 Modified
9	45	100.0	701	4	AAE05117	Aae05117 Modified
10	45	100.0	701	5	AAB47919	Aab47919 CEA agoni
11	45	100.0	701	5	AAB47918	Aab47918 CEA agoni
12	43	95.6	9	2	AAY09527	Aay09527 Carcinoem
13	40	88.9	9	2	AAW39723	Aaw39723 Human car
14	40	88.9	9	2	AAW70045	Aaw70045 CEA deriv
15	40	88.9	9	2	AAW77134	Aaw77134 CEA synth
16	40	88.9	9	2	AAY47655	Aay47655 Immunogen
17	40	88.9	9	2	AAY09525	Aay09525 Carcinoem
18	40	88.9	9	3	AAB13749	Aab13749 Peptide f
19	40	88.9	9	4	AAE02673	Aae02673 Human CEA
20	40	88.9	9	4	AAE00463	Aae00463 Human tum
21	40	88.9	9	4	AAE05123	Aae05123 Carcinoem
22	40	88.9	9	4	AAB82776	Aab82776 Carcinoem
23	40	88.9	9	5	ABG79073	Abg79073 Human CEA
24	40	88.9	9	5	AAE26805	Aae26805 Human HLA
25	40	88.9	9	5	AAU95893	Aau95893 Immunogen
26	40	88.9	9	5	AAE19088	Aae19088 HLA-A24 r
27	40	88.9	9	6	ABR56428	Abr56428 CEA epito
28	40	88.9	9	6	ABP98779	Abp98779 CAE pepti
29	40	88.9	9	6	ABR44529	Abr44529 CEA epito
30	40	88.9	9	7	ADD84715	Add84715 Human car
31	40	88.9	9	7	AAO24210	Aao24210 Human tum
32	40	88.9	9	8	ADG20333	Adg20333 Antigenic
33	40	88.9	9	8	ADJ36382	Adj36382 CEA epito
34	40	88.9	9	8	ADM12344	Adm12344 MHC class
35	40	88.9	9	8	ADM12341	Adm12341 MHC class
36	40	88.9	9	8	ADM72999	Adm72999 Human CEA
37	40	88.9	9	8	ADL46188	Adl46188 Human CAP
38	40	88.9	9	8	ADO38561	Ado38561 Carcinoem
39	40	88.9	9	8	ADO38564	Ado38564 Carcinoem
40	40	88.9	10	2	AAY46555	Aay46555 Immunogen
41	40	88.9	10	5	AAU11587	Aau11587 Human car
42	40	88.9	10	6	ABR83489	Abr83489 Human car
43	40	88.9	10	8	ADM72998	Adm72998 Human CEA
44	40	88.9	10	8	ADP80031	Adp80031 Human HLA
45	40	88.9	14	4	AAB88124	Aab88124 CD66 pept

ALIGNMENTS

RESULT 1
AAY09526
ID AAY09526 standard; peptide; 9 AA.
XX
AC AAY09526;
XX
DT 20-JUL-1999 (first entry)
XX
DE Carcinoembryonic antigen peptide agonist SEQ ID NO:2.
XX
KW Carcinoembryonic antigen; CEA; human; agonist; antagonist;
KW immune response; carcinoma; gastrointestinal; breast; pancreatic;
KW bladder; ovarian; lung; prostatic; T cell proliferation; cancer;
KW adoptive transfer therapy; autoimmune reaction; immunotherapy.
XX
OS Homo sapiens.
OS Synthetic.
XX
PN WO9919478-A1.
XX
PD 22-APR-1999.
XX
PF 22-SEP-1998; 98WO-US019794.
XX
PR 10-OCT-1997; 97US-0061589P.
XX
PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
XX
PI Schlom J, Barzaga E, Zaremba S;
XX
DR WPI; 1999-326544/27.
XX
PT Peptide agonists and antagonists of carcinoembryonal antigen.
XX
PS Claim 5; Page 53; 72pp; English.
XX
CC The present invention describes peptides (A) that comprise agonists (Ia)
CC or antagonists (Ib) of human carcinoembryonal antigen (CEA). (Ia) are
CC used in vaccines to kill or inhibit carcinoma cells that express CEA or
CC its epitopes, particularly for treating gastrointestinal, breast,
CC pancreatic, bladder, ovarian, lung or prostatic carcinoma. They can also
CC be used to proliferate T cells, e.g. from vaccinated subjects, for use in
CC adoptive transfer therapy. (Ib) are used to inhibit CEA-specific immune
CC responses, e.g. in vaccinated subjects, to prevent an autoimmune reaction
CC to cancer immunotherapy (i.e. to prevent attack on normal but CEA-
CC expressing cells). (Ia) are more active than native sequence (I) and
CC generate a highly specific and systemic anti-CEA response. Cytotoxic T
CC cells generated recognize both (Ia) and native CEA epitopes. The present

```
CC   sequence represents a specifically claimed example of (Ia)
XX
SQ   Sequence 9 AA;

  Query Match             100.0%;  Score 45;  DB 2;  Length 9;
  Best Local Similarity   100.0%;  Pred. No. 1.8e+06;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 YLSGADLNL 9
              |||||||||
Db          1 YLSGADLNL 9


RESULT 2
AAB13750
ID   AAB13750 standard; peptide; 9 AA.
XX
AC   AAB13750;
XX
DT   02-FEB-2001  (first entry)
XX
DE   Peptide fragment # 2 from human CEA.
XX
KW   Human; T-cell; immune response; antigen; epitope; B7 family molecule;
KW   Leukocyte function-associated antigen-3; LFA-3;
KW   Intercellular adhesion molecule-1; ICAM-1; vaccine; immunotherapy;
KW   colon polyp; Crohn's disease; ulcerative colitis; breast lesion; tumour;
KW   CEA.
XX
OS   Homo sapiens.
XX
PN   WO200034494-A1.
XX
PD   15-JUN-2000.
XX
PF   12-NOV-1999;   99WO-US026866.
XX
PR   09-DEC-1998;   98US-0111582P.
XX
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICES.
PA   (THER-) THERION BIOLOGICS CORP.
XX
PI   Schlom J,  Hodge J,  Panicali D;
XX
DR   WPI; 2000-431307/37.
XX
PT   Novel recombinant vector useful as immunogens and vaccines for
PT   stimulating and enhancing immunological responses to target cells and
PT   antigens expresses multiple co-stimulatory molecules such as B7-1, LFA-3,
PT   ICAM-1.
XX
PS   Claim 18; Page 35; 188pp; English.
XX
CC   Costimulatory molecules have important roles in T-cell activation and
CC   therefore the immune response. The present invention relates to
CC   recombinant vectors which comprise of foreign nucleic acid sequences
CC   encoding at least three costimulatory molecules: a B7 family molecule,
CC   Leukocyte function-associated antigen-3 (LFA-3, human CD58) and
CC   Intercellular adhesion molecule-1 (ICAM-1, CD54) and optionally a foreign
CC   gene encoding a target antigen or immunological epitope. The present
CC   sequence is one such target antigen used in the present invention. The
CC   present sequence is a tumour-associated antigen. The vector of the
CC   present invention would be useful for providing an enhanced immune
CC   response to the present target antigen. The vector of the present
CC   invention may therefore be useful in immunotherapy for treating or
CC   preventing diseases caused by viruses, bacteria, protozoans, parasites,
CC   premalignant cells and tumour cells. The recombinant vector can be used
CC   to treat or prevent preneoplastic or hyperplastic states such as colon
CC   polyps, Crohn's disease, ulcerative colitis and breast lesions
XX
SQ   Sequence 9 AA;

  Query Match             100.0%;  Score 45;  DB 3;  Length 9;
  Best Local Similarity   100.0%;  Pred. No. 1.8e+06;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 YLSGADLNL 9
              |||||||||
Db          1 YLSGADLNL 9


RESULT 3
AAB97818
ID   AAB97818 standard; peptide; 9 AA.
XX
AC   AAB97818;
XX
DT   08-AUG-2001  (first entry)
XX
DE   Carcinoembryonic antigen (CEA) modified antigen SEQ ID NO:113.
XX
KW   Virus; adenovirus; poxvirus; alphavirus; immune response; gp100;
KW   tumour antigen; CEA; carcinoembryonic antigen; immunostimulant;
KW   cytostatic; immunotherapy; interferon-gamma; IFN-gamma; cancer.
XX
OS   Unidentified.
XX
PN   WO200130382-A1.
XX
PD   03-MAY-2001.
XX
PF   20-OCT-2000; 2000WO-CA001253.
XX
PR   22-OCT-1999;   99US-0160879P.
PR   07-AUG-2000; 2000US-0223325P.
XX
PA   (AVET ) AVENTIS PASTEUR LTD.
XX
PI   Berinstein N,  Tartaglia J,  Moingeon P,  Barber B;
XX
DR   WPI; 2001-308587/32.
XX
PT   Inducing immune response to tumor antigen, useful in immunotherapy of
PT   cancer, by administering the antigen to a lymphatic site.
XX
PS   Claim 19; Page 9; 60pp; English.
XX
CC   The present invention describes a method for inducing an immune response,
CC   in an animal, to a tumour antigen (Ag) comprising administering Ag, or
CC   nucleic acid (I) that encodes it, to a lymphatic site. Cynomolgus monkeys
CC   (Macaca fascicularis) were injected with a modified form of gp100 antigen
CC   (a) into the left inguinal lymph node or (b) subcutaneously. Both animals
CC   of (a) developed a cell-mediated response (indicated by production of
CC   interferon-gamma from T lymphocytes when exposed to gp100 peptides), but
CC   only 2 of 4 animals of (b) did so. Also animals in (a) produced a far
CC   greater antibody response to gp100. The method is used in immunotherapy
CC   of a wide range of cancers through induction of a specific immune
CC   response (humoral and cellular) against the tumour antigens. When
CC   administered to a lymphatic site, Ag (or (I)) induces a stronger immune
CC   response than administration by other routes and may also break tolerance
CC   to Ag. AAB97708 and AAB97709 represent gp100 epitopes; AAB97710 to
CC   AAB97815 represent peptides derived from gp100 which stimulate interferon
CC   (IFN)-gamma production; AAH20120 encodes the modified gp100 protein given
CC   in AAB97816; AAH20121 encodes the modified carcinoembryonic antigen (CEA)
CC   protein given in AAB97817; and AAB97818 represents a CEA modified antigen
CC   peptide, all of which are used in the exemplification of the present
CC   invention
XX
SQ   Sequence 9 AA;

  Query Match             100.0%;  Score 45;  DB 4;  Length 9;
  Best Local Similarity   100.0%;  Pred. No. 1.8e+06;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 YLSGADLNL 9
              |||||||||
```

```
Db          1 YLSGADLNL 9


RESULT 4
AAE05124
ID   AAE05124 standard; peptide; 9 AA.
XX
AC   AAE05124;
XX
DT   18-SEP-2001  (first entry)
XX
DE   Modified carcinoembryonic antigen (CEA) peptide, CAP-6D.
XX
KW   Tumour-associated antigen; TAA; cytostatic; vaccine; gene therapy;
KW   immune response; tetanus toxoid; TT; diphtheria toxoid; DT; prophylactic;
KW   cancer; therapeutic; carcinoembryonic antigen; CEA.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 6
FT                   /note= "Wild type Asn substituted with Asp"
XX
PN   WO200149317-A2.
XX
PD   12-JUL-2001.
XX
PF   05-JAN-2001; 2001WO-CA000005.
XX
PR   05-JAN-2000; 2000US-0174587P.
XX
PA   (AVET ) AVENTIS PASTEUR LTD.
XX
PI   Emtage P,  Barber BH,  Sambhara S,  Sia CDY;
XX
DR   WPI; 2001-441790/47.
XX
PT   Enhancing immune response to antigen such as tumor antigen for treating
PT   cancer in an animal involves administering an inducing agent to the
PT   animal followed by administering inducing agent-antigen mixture.
XX
PS   Example 2; Page 31; 62pp; English.
XX
CC   The invention relates to a method of enhancing an immune response against
CC   tumour-associated antigens (TAAs), such as GP100 and carcinoembryonic
CC   antigen (CEA) in an animal. The method involves priming of the animal
CC   with an inducing agent such as tetanus toxoid (TT) or diphtheria toxoid
CC   (DT), subsequently followed by administration of an inducing agent-
CC   antigen mixture. The method provides the enhancement or augmentation of
CC   the immune response to the antigen and/or improves a vaccination protocol
CC   by allowing use of less antigen. The immunisation of the animal with
CC   tumour-associated antigen is useful for the prophylactic or therapeutic
CC   treatment of cancer. The present sequence is modified carcinoembryonic
CC   antigen (CEA) peptide fragment related to the invention
XX
SQ   Sequence 9 AA;

  Query Match             100.0%;  Score 45;  DB 4;  Length 9;
  Best Local Similarity   100.0%;  Pred. No. 1.8e+06;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 YLSGADLNL 9
              |||||||||
Db          1 YLSGADLNL 9


RESULT 5
AAB47917
ID   AAB47917 standard; peptide; 9 AA.
XX
AC   AAB47917;
XX
DT   16-MAY-2002  (first entry)
XX
DE   Modified CEA epitope, CEA(6D).
XX
KW   CAP-1; epitope; carcinoembryonic antigen; CEA; agonist; immune response;
KW   carcinoma; gastrointestinal; breast; pancreatic; bladder; ovarian; lung;
KW   prostate; cancer.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 6
FT                   /label= N6D
XX
PN   WO200210379-A2.
XX
PD   07-FEB-2002.
XX
PF   27-JUL-2001; 2001WO-CA001092.
XX
PR   31-JUL-2000; 2000US-0222043P.
XX
PA   (AVET ) AVENTIS PASTEUR LTD.
PA   (THER-) THERION BIOLOGICS.
PA   (USSH ) US NAT CANCER INST.
XX
PI   Berinstein N,  Tartaglia J,  Tine JA,  Panicali DL,  Gritz L;
PI   Schlom J;
XX
DR   WPI; 2002-206189/26.
XX
PT   Carcinoembryonic antigen agonist polypeptide for inducing an immune
PT   response in animal against antigen and for inhibiting an epitope antigen
PT   expressing carcinoma cell, comprises a modified antigen epitope.
XX
PS   Claim 1; Page 38; 69pp; English.
XX
CC   This sequence represents a modified CAP-1 epitope of carcinoembryonic
CC   antigen (CEA) which was used as part of the CEA agonist polypeptide of
CC   the invention. The modification of position 6 of this peptide from Asp to
CC   Asn increases its immunogenicity. The CEA agonist polypeptide of the
CC   invention, or DNA encoding it, are useful for: (i) inducing an immune
CC   response in an animal directed against a CEA protein or fragment, CEA
CC   agonist, a CEA epitope, a modified CEA epitope, cells expressing or
CC   binding a CEA protein or fragment; and (ii) inhibiting a CEA epitope
CC   expressing carcinoma cell, which is a gastrointestinal, breast,
CC   pancreatic, bladder, ovarian, lung or prostate carcinoma cell in a
CC   patient, hence is useful for manufacture of a medicament for the
CC   treatment of cancer
XX
SQ   Sequence 9 AA;

  Query Match             100.0%;  Score 45;  DB 5;  Length 9;
  Best Local Similarity   100.0%;  Pred. No. 1.8e+06;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 YLSGADLNL 9
              |||||||||
Db          1 YLSGADLNL 9


RESULT 6
AAE19089
ID   AAE19089 standard; peptide; 9 AA.
XX
AC   AAE19089;
XX
DT   21-MAY-2002  (first entry)
XX
DE   HLA-A24 restricted target antigen CEA immunological epitope #3.
XX
KW   Human leukocyte antigen; HLA; pharmaceutical composition; target antigen;
KW   immunological epitope; replication-defective virus; RDV; immune response;
KW   chemotherapy; granulocyte-monocyte-colony stimulating factor; cytostatic;
```

KW GM-CSF; MHC; major histocompatibility complex; tumour; head; pancreatic;
KW neck; breast; prostate; colorectal; melanoma; myeloidysplastic syndrome;
KW metastatic breast skin lesion; corticosteroid therapy; erythropoietin;
KW cytopenia; neutropenia; vaccine; immunostimulant.
XX
OS Homo sapiens.
XX
PN WO200195919-A2.
XX
PD 20-DEC-2001.
XX
PF 15-JUN-2001; 2001WO-US019201.
XX
PR 15-JUN-2000; 2000US-0211717P.
XX
PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
PA (THER-) THERION BIOLOGICS CORP.
XX
PI Schlom J, Greiner JW, Kass E, Panicali D;
XX
DR WPI; 2002-205852/26.
XX
PT Composition for enhancing immune responses, particularly anti-tumor
PT responses and treating neutropenia, cytopenia, comprises replication-
PT defective virus encoding granulocyte-monocyte-colony stimulating factor.
XX
PS Claim 9; Page 15; 118pp; English.
XX
CC The present invention relates to a pharmaceutical composition comprising
CC a replication-defective virus (RDV) encoding granulocyte-monocyte-colony
CC stimulating factor (GM-CSF). The invention is useful for enhancing cell-
CC mediated or humoral immune response in an individual, by enhancing
CC migration of APC expressing CD11c^+/I-Ab^+, major histocompatibility
CC complex (MHC) class II, at an injection site, regional lymph node at a
CC tumour site, APC proliferation or function, CD4^+T or CD8^+T cell
CC activation, interleukin (IL)-2, interferon (IFN)-gamma or tumour necrosis
CC factor (TNF)-alpha production or their combinations. The composition
CC enhances an antigen-specific T-cell response in an individual to a target
CC antigen or its immunological epitope and an anti-tumour response in an
CC individual with a head tumour, neck, breast, pancreatic, prostate,
CC colorectal or metastatic tumour or melanoma, or metastatic breast skin
CC lesion. The invention is further useful for treating neutropenia
CC resulting from chemotherapy, corticosteroid therapy, irradiation or an
CC infection, by raising the neutrophil count to normal levels and for
CC treating cytopenias in patients with myeloidysplastic syndrome in
CC combination with erythropoietin, by increasing neutrophil count and
CC erythroid precursors. The composition enhances immune response to
CC vaccines such as DPT, Td, DtaP, Hib, DtaP-Hib, MMR, Hepatitis A,
CC hepatitis B, lyme's disease, influenza, tetravalent meningococcal
CC polysaccharide, pneumococcal polysaccharide, anthrax, cholera, plague,
CC yellow fever and Bacillus Calmette-Guerin vaccine. The present sequence
CC is human leukocyte antigen (HLA)-restricted target tumour antigen
CC immunological epitope
XX
SQ Sequence 9 AA;

Query Match 100.0%; Score 45; DB 5; Length 9;
Best Local Similarity 100.0%; Pred. No. 1.8e+06;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 YLSGADLNL 9
|||||||||
Db 1 YLSGADLNL 9

RESULT 7
ADN63713
ID ADN63713 standard; peptide; 9 AA.
XX
AC ADN63713;
XX
DT 01-JUL-2004 (first entry)
XX
DE HLA binding peptide #313.
XX
KW cytostatic; hepatotropic; virucide; antiinflammatory; anti-HIV;
KW gene therapy; vaccine; HLA binding peptide; HTL epitope; liposome;
KW prostate specific antigen; prostate specific membrane antigen;
KW hepatitis B virus antigen; hepatitis C virus antigen;
KW malignant melanoma antigen; MAGE; Epstein Barr virus; cancer;
KW prostate cancer; AIDS; renal carcinoma; cervical carcinoma; lymphoma;
KW chondyloma acuminatum.
XX
OS Unidentified.
XX
PN WO2004031211-A2.
XX
PD 15-APR-2004.
XX
PF 03-OCT-2003; 2003WO-US031308.
XX
PR 03-OCT-2002; 2002US-0416207P.
PR 08-OCT-2002; 2002US-0417269P.
XX
PA (EPIM-) EPIMMUNE INC.
XX
PI Sidney J, Southwood S, Sette A;
XX
DR WPI; 2004-347953/32.
XX
PT New composition of peptides and nucleic acids capable of binding Major
PT Histocompatibility Complex molecules, useful for diagnosing, preventing
PT or treating viral infections or cancer, such as prostate cancer,
PT hepatitis B or AIDS.
XX
PS Claim 1; SEQ ID NO 313; 186pp; English.
XX
CC The invention relates to a novel composition comprising one or more
CC peptides or nucleic acids encoding an HLA binding peptide. The
CC composition further comprises an HTL epitope. It also comprises a spacer
CC molecule, a carrier, an MHC targeting sequence or a lipid. The peptides
CC are incorporated as part of a liposome. The peptide is from an antigen
CC selected from prostate specific antigen (PSA), prostate specific membrane
CC antigen (PSM), hepatitis B virus (HBV) antigen, hepatitis C virus (HCV)
CC antigen, malignant melanoma antigen (MAGE), Epstein Barr virus, human
CC immunodeficiency type-1 (HIV-1), human immunodeficiency type-2 (HIV-2),
CC Papilloma virus, Lassa virus, Mycobacterium tuberculosis (MT), p53,
CC murine p53 (mp53), CEA, HER2/neu, and tyrosine kinase related protein
CC (TKP). The composition is useful for preventing or treating viral
CC infections or cancer, such as prostate cancer, hepatitis B, hepatitis C,
CC AIDS, renal carcinoma, cervical carcinoma, lymphoma, CMV or chondyloma
CC acuminatum. The composition is also be used for diagnosing such diseases.
CC This sequence represents a peptide of the invention.
XX
SQ Sequence 9 AA;

Query Match 100.0%; Score 45; DB 8; Length 9;
Best Local Similarity 100.0%; Pred. No. 1.8e+06;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 YLSGADLNL 9
|||||||||
Db 1 YLSGADLNL 9

RESULT 8
AAB97817
ID AAB97817 standard; protein; 701 AA.
XX
AC AAB97817;
XX
DT 08-AUG-2001 (first entry)
XX
DE Modified Carcinoembryonic antigen (CEA) protein SEQ ID NO:112.
XX
KW Virus; adenovirus; poxvirus; alphavirus; immune response; gp100;

KW tumour antigen; CEA; carcinoembryonic antigen; immunostimulant;
KW cytostatic; immunotherapy; interferon-gamma; IFN-gamma; cancer.
XX
OS Unidentified.
XX
PN WO200130382-A1.
XX
PD 03-MAY-2001.
XX
PF 20-OCT-2000; 2000WO-CA001253.
XX
PR 22-OCT-1999; 99US-0160879P.
PR 07-AUG-2000; 2000US-0223325P.
XX
PA (AVET) AVENTIS PASTEUR LTD.
XX
PI Berinstein N, Tartaglia J, Moingeon P, Barber B;
XX
DR WPI; 2001-308587/32.
DR N-PSDB; AAH20121.
XX
PT Inducing immune response to tumor antigen, useful in immunotherapy of
PT cancer, by administering the antigen to a lymphatic site.
XX
PS Claim 19; Fig 8; 60pp; English.
XX
CC The present invention describes a method for inducing an immune response,
CC in an animal, to a tumour antigen (Ag) comprising administering Ag, or
CC nucleic acid (I) that encodes it, to a lymphatic site. Cynomolgus monkeys
CC (Macaca fascicularis) were injected with a modified form of gp100 antigen
CC (a) into the left inguinal lymph node or (b) subcutaneously. Both animals
CC of (a) developed a cell-mediated response (indicated by production of
CC interferon-gamma from T lymphocytes when exposed to gp100 peptides), but
CC only 2 of 4 animals of (b) did so. Also animals in (a) produced a far
CC greater antibody response to gp100. The method is used in immunotherapy
CC of a wide range of cancers through induction of a specific immune
CC response (humoral and cellular) against the tumour antigens. When
CC administered to a lymphatic site, Ag (or (I)) induces a stronger immune
CC response than administration by other routes and may also break tolerance
CC to Ag. AAB97708 and AAB97709 represent gp100 epitopes; AAB97710 to
CC AAB97815 represent peptides derived from gp100 which stimulate interferon
CC (IFN)-gamma production; AAH20120 encodes the modified gp100 protein given
CC in AAB97816; AAH20121 encodes the modified carcinoembryonic antigen (CEA)
CC protein given in AAB97817; and AAB97818 represents a CEA modified antigen
CC peptide, all of which are used in the exemplification of the present
CC invention
XX
SQ Sequence 701 AA;

Query Match 100.0%; Score 45; DB 4; Length 701;
Best Local Similarity 100.0%; Pred. No. 2.8;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 YLSGADLNL 9
|||||||||
Db 604 YLSGADLNL 612

RESULT 9
AAE05117
ID AAE05117 standard; protein; 701 AA.
XX
AC AAE05117;
XX
DT 18-SEP-2001 (first entry)
XX
DE Modified carcinoembryonic antigen (CEA).
XX
KW Tumour-associated antigen; TAA; cytostatic; vaccine; gene therapy;
KW immune response; tetanus toxoid; TT; diphtheria toxoid; DT; prophylactic;
KW cancer; therapeutic; carcinoembryonic antigen; CEA.
XX
OS Synthetic.
XX
PN WO200149317-A2.
XX
PD 12-JUL-2001.
XX
PF 05-JAN-2001; 2001WO-CA000005.
XX
PR 05-JAN-2000; 2000US-0174587P.
XX
PA (AVET) AVENTIS PASTEUR LTD.
XX
PI Emtage P, Barber BH, Sambhara S, Sia CDY;
XX
DR WPI; 2001-441790/47.
DR N-PSDB; AAD07347.
XX
PT Enhancing immune response to antigen such as tumor antigen for treating
PT cancer in an animal involves administering an inducing agent to the
PT animal followed by administering inducing agent-antigen mixture.
XX
PS Claim 9; Fig 3; 62pp; English.
XX
CC The invention relates to a method of enhancing an immune response against
CC tumour-associated antigens (TAAs), such as GP100 and carcinoembryonic
CC antigen (CEA) in an animal. The method involves priming of the animal
CC with an inducing agent such as tetanus toxoid (TT) or diphtheria toxoid
CC (DT), subsequently followed by administration of an inducing agent-
CC antigen mixture. The method provides the enhancement or augmentation of
CC the immune response to the antigen and/or improves a vaccination protocol
CC by allowing use of less antigen. The immunisation of the animal with
CC tumour-associated antigen is useful for the prophylactic or therapeutic
CC treatment of cancer. The present sequence is modified carcinoembryonic
CC antigen (CEA) related to the invention
XX
SQ Sequence 701 AA;

Query Match 100.0%; Score 45; DB 4; Length 701;
Best Local Similarity 100.0%; Pred. No. 2.8;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 YLSGADLNL 9
|||||||||
Db 604 YLSGADLNL 612

RESULT 10
AAB47919
ID AAB47919 standard; protein; 701 AA.
XX
AC AAB47919;
XX
DT 16-MAY-2002 (first entry)
XX
DE CEA agonist #2.
XX
KW CAP-1; epitope; carcinoembryonic antigen; CEA; agonist; immune response;
KW carcinoma; gastrointestinal; breast; pancreatic; bladder; ovarian; lung;
KW prostate; cancer.
XX
OS Synthetic.
XX
PN WO200210379-A2.
XX
PD 07-FEB-2002.
XX
PF 27-JUL-2001; 2001WO-CA001092.
XX
PR 31-JUL-2000; 2000US-0222043P.
XX
PA (AVET) AVENTIS PASTEUR LTD.
PA (THER-) THERION BIOLOGICS.
PA (USSH) US NAT CANCER INST.
XX

PI Berinstein N, Tartaglia J, Tine JA, Panicali DL, Gritz L;
PI Schlom J;
XX
DR WPI; 2002-206189/26.
XX
PT Carcinoembryonic antigen agonist polypeptide for inducing an immune
PT response in animal against antigen and for inhibiting an epitope antigen
PT expressing carcinoma cell, comprises a modified antigen epitope.
XX
PS Claim 2; Page 64-66; 69pp; English.
XX
CC This sequence shows the carcinoembryonic antigen (CEA) agonist
CC polypeptide of the invention. This sequence represents the sequence given
CC in the Seq ID listing in the specification, and is not directly encoded
CC by the coding sequence given in AAI72497 . The CEA agonist contains a
CC modified CAP-1 epitope of CEA, in which position 6 is modified from Asp
CC to Asn to increase its immunogenicity. The CEA agonist polypeptide of the
CC invention, or DNA encoding it, are useful for: (i) inducing an immune
CC response in an animal directed against a CEA protein or fragment, CEA
CC agonist, a CEA epitope, a modified CEA epitope, cells expressing or
CC binding a CEA protein or fragment; and (ii) inhibiting a CEA epitope
CC expressing carcinoma cell, which is a gastrointestinal, breast,
CC pancreatic, bladder, ovarian, lung or prostate carcinoma cell in a
CC patient, hence is useful for manufacture of a medicament for the
CC treatment of cancer
XX
SQ Sequence 701 AA;

Query Match 100.0%; Score 45; DB 5; Length 701;
Best Local Similarity 100.0%; Pred. No. 2.8;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 YLSGADLNL 9
|||||||||
Db 604 YLSGADLNL 612

RESULT 11
AAB47918
ID AAB47918 standard; protein; 701 AA.
XX
AC AAB47918;
XX
DT 16-MAY-2002 (first entry)
XX
DE CEA agonist #1.
XX
KW CAP-1; epitope; carcinoembryonic antigen; CEA; agonist; immune response;
KW carcinoma; gastrointestinal; breast; pancreatic; bladder; ovarian; lung;
KW prostate; cancer.
XX
OS Synthetic.
XX
PN WO200210379-A2.
XX
PD 07-FEB-2002.
XX
PF 27-JUL-2001; 2001WO-CA001092.
XX
PR 31-JUL-2000; 2000US-0222043P.
XX
PA (AVET) AVENTIS PASTEUR LTD.
PA (THER-) THERION BIOLOGICS.
PA (USSH) US NAT CANCER INST.
XX
PI Berinstein N, Tartaglia J, Tine JA, Panicali DL, Gritz L;
PI Schlom J;
XX
DR WPI; 2002-206189/26.
DR N-PSDB; AAI72489.
XX
PT Carcinoembryonic antigen agonist polypeptide for inducing an immune
PT response in animal against antigen and for inhibiting an epitope antigen
PT expressing carcinoma cell, comprises a modified antigen epitope.
XX
PS Claim 2; Fig 1; 69pp; English.
XX
CC This sequence shows the carcinoembryonic antigen (CEA) agonist
CC polypeptide of the invention. This sequence represents the sequence given
CC in the figures in the specification, and is directly encoded by the
CC coding sequence given in AAI72489. The CEA agonist contains a modified
CC CAP-1 epitope of CEA, in which position 6 is modified from Asp to Asn to
CC increase its immunogenicity. The CEA agonist polypeptide of the
CC invention, or DNA encoding it, are useful for: (i) inducing an immune
CC response in an animal directed against a CEA protein or fragment, CEA
CC agonist, a CEA epitope, a modified CEA epitope, cells expressing or
CC binding a CEA protein or fragment; and (ii) inhibiting a CEA epitope
CC expressing carcinoma cell, which is a gastrointestinal, breast,
CC pancreatic, bladder, ovarian, lung or prostate carcinoma cell in a
CC patient, hence is useful for manufacture of a medicament for the
CC treatment of cancer
XX
SQ Sequence 701 AA;

Query Match 100.0%; Score 45; DB 5; Length 701;
Best Local Similarity 100.0%; Pred. No. 2.8;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 YLSGADLNL 9
|||||||||
Db 604 YLSGADLNL 612

RESULT 12
AAY09527
ID AAY09527 standard; peptide; 9 AA.
XX
AC AAY09527;
XX
DT 20-JUL-1999 (first entry)
XX
DE Carcinoembryonic antigen peptide agonist SEQ ID NO:3.
XX
KW Carcinoembryonic antigen; CEA; human; agonist; antagonist;
KW immune response; carcinoma; gastrointestinal; breast; pancreatic;
KW bladder; ovarian; lung; prostatic; T cell proliferation; cancer;
KW adoptive transfer therapy; autoimmune reaction; immunotherapy.
XX
OS Homo sapiens.
OS Synthetic.
XX
PN WO9919478-A1.
XX
PD 22-APR-1999.
XX
PF 22-SEP-1998; 98WO-US019794.
XX
PR 10-OCT-1997; 97US-0061589P.
XX
PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
XX
PI Schlom J, Barzaga E, Zaremba S;
XX
DR WPI; 1999-326544/27.
XX
PT Peptide agonists and antagonists of carcinoembryonal antigen.
XX
PS Claim 5; Page 53; 72pp; English.
XX
CC The present invention describes peptides (A) that comprise agonists (Ia)
CC or antagonists (Ib) of human carcinoembryonal antigen (CEA). (Ia) are
CC used in vaccines to kill or inhibit carcinoma cells that express CEA or
CC its epitopes, particularly for treating gastrointestinal, breast,
CC pancreatic, bladder, ovarian, lung or prostatic carcinoma. They can also
CC be used to proliferate T cells, e.g. from vaccinated subjects, for use in
CC adoptive transfer therapy. (Ib) are used to inhibit CEA-specific immune

CC responses, e.g. in vaccinated subjects, to prevent an autoimmune reaction
CC to cancer immunotherapy (i.e. to prevent attack on normal but CEA-
CC expressing cells). (Ia) are more active than native sequence (I) and
CC generate a highly specific and systemic anti-CEA response. Cytotoxic T
CC cells generated recognize both (Ia) and native CEA epitopes. The present
CC sequence represents a specifically claimed example of (Ia)
XX
SQ Sequence 9 AA;

Query Match 95.6%; Score 43; DB 2; Length 9;
Best Local Similarity 88.9%; Pred. No. 1.8e+06;
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 YLSGADLNL 9
||||||:||
Db 1 YLSGADINL 9

RESULT 13
AAW39723
ID AAW39723 standard; peptide; 9 AA.
XX
AC AAW39723;
XX
DT 11-JUN-1998 (first entry)
XX
DE Human carcina-embryonic antigen (CEA) peptide (pos. 571-579).
XX
KW T cell epitope; immune response; human leukocyte antigen; HLA Class I;
KW vaccine; immunogenic; major histocompatablilty complex; MHC; B cell;
KW disease; anti-tumour; anti-viral.
XX
OS Homo sapiens.
XX
PN WO9741440-A1.
XX
PD 06-NOV-1997.
XX
PF 28-APR-1997; 97WO-NL000229.
XX
PR 26-APR-1996; 96EP-00201145.
PR 23-DEC-1996; 96EP-00203670.
XX
PA (UYLE-) RIJKSUNIV LEIDEN.
PA (SCIS-) SCI SEED CAPITAL INVESTMENTS BV.
XX
PI Van Der Burg SH, Kast WM, Toes REM, Offringa R, Melief CJM;
XX
DR WPI; 1997-549891/50.
XX
PT Method of selecting T cell peptide epitope(s) - by measuring the
PT stability of HLA class I-peptide complexes on intact B cells.
XX
PS Example 3; Page 85; 109pp; English.
XX
CC Peptides AAW39430-W39734 are used in a novel method for the selection of
CC immunogenic T-cell peptide epitopes present in polypeptide antigens. The
CC method involves the identification of peptide sequences capable of
CC binding to an HLA (human leukocyte antigen) class I molecule and
CC measuring the binding of this epitope peptide to the HLA class I peptide.
CC The stability of binding of the peptide and MHC (major histocompatablilty
CC complex) class I molecule is measured on intact human B cells carrying
CC the MHC molecule at their cell surfaces. The method can be used to select
CC peptide epitopes for generating vaccines against a disease associated
CC with the polypeptide, e.g. cancers or AIDS. The peptide epitopes are
CC especially T-cell peptide epitopes with strong anti-tumour and anti-viral
CC immune responses. Peptide AAW39723 is derived from the human carcino-
CC embryonic antigen (CEA) and has the ability to bind to the human MHC
CC Class I allele HLA-A2.1
XX
SQ Sequence 9 AA;

Query Match 88.9%; Score 40; DB 2; Length 9;
Best Local Similarity 88.9%; Pred. No. 1.8e+06;
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 YLSGADLNL 9
|||||:|||
Db 1 YLSGANLNL 9

RESULT 14
AAW70045
ID AAW70045 standard; peptide; 9 AA.
XX
AC AAW70045;
XX
DT 22-OCT-1998 (first entry)
XX
DE CEA derived HLA-A2.1 binding peptide 2 (residues 605-613).
XX
KW Cytotoxic T lymphocyte; CTL; major histocompatibility complex; MHC;
KW human leukocyte antigen; HLA; tumour associated antigen; cancer;
KW antigen presenting cell; APC; immunogenic peptide; immune disorder;
KW viral infection; AIDS; hepatitis; bacterial infection; malaria; CEA;
KW fungal infection; tuberculosis; melanoma; carcinoembryonic antigen.
XX
OS Synthetic.
OS Homo sapiens.
XX
PN WO9833888-A1.
XX
PD 06-AUG-1998.
XX
PF 30-JAN-1998; 98WO-US001959.
XX
PR 31-JAN-1997; 97US-0036696P.
XX
PA (EPIM-) EPIMMUNE INC.
XX
PI Tsai V, Southwood S, Sidney J, Sette A, Celis E;
XX
DR WPI; 1998-437445/37.
XX
PT Production of antigen-specific cytotoxic T cells - by incubating
PT immunogenic peptide(s) from antigen that binds class I major
PT histocompatibility complex molecules with pre-treated antigen presenting
PT cells.
XX
PS Example 6; Page 75; 104pp; English.
XX
CC Sequences shown in AAW70044 to AAW70052 represent peptides derived from
CC carcinoembryonic antigen (CEA). The peptides can bind to a human
CC leukocyte antigen (HLA), HLA-A2.1 and are used to exemplify the method of
CC invention of producing antigen-specific cytotoxic T cells (CTLs) in
CC vitro. The method comprises contacting immunogenic peptides from an
CC antigen that binds class I major histocompatibility complex (MHC)
CC molecules with antigen presenting cells (APCs) pretreated with
CC pretreatment growth factors, and incubating the APCs with purified CD8
CC cells in the presence of at least 2 incubation growth factors, thereby
CC producing antigen-specific CTLs. A method for specifically killing target
CC cells in a human patient is also provided which comprises obtaining a
CC fluid sample containing CTLs from a patient, contacting the cytotoxic T
CC cells with APCs pretreated with pre-treatment growth factors, where the
CC APCs comprise class I MHC molecules. The pretreated APCs are incubated
CC with the cytotoxic growth factors, thereby producing activated CTLs which
CC are contacted with a carrier to form a composition. The composition can
CC then be administered to the patient. The activated CTLs can be used for
CC treating cancers, immune disorders, viral infections, AIDS, hepatitis,
CC bacterial infection, fungal infection, malaria or tuberculosis
XX
SQ Sequence 9 AA;

Query Match 88.9%; Score 40; DB 2; Length 9;
Best Local Similarity 88.9%; Pred. No. 1.8e+06;
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 YLSGADLNL 9
              |||||:|||
Db          1 YLSGANLNL 9
```

RESULT 15
AAW77134
ID AAW77134 standard; peptide; 9 AA.
XX
AC AAW77134;
XX
DT 16-NOV-1998 (first entry)
XX
DE CEA synthetic peptide epitope 1.
XX
KW Tyrosinase; tyrosinase cytotoxic lymphocyte response;
KW cytotoxic T lymphocyte; cysteine-depleted; melanoma.
XX
OS Synthetic.
XX
PN WO9833810-A2.
XX
PD 06-AUG-1998.
XX
PF 29-JAN-1998; 98WO-US001592.
XX
PR 30-JAN-1997; 97US-0037781P.
XX
PA (UYVI-) UNIV VIRGINIA PATENT FOUND.
XX
PI Slingluff CL, Hunt DF, Engelhard VH, Kittlesen D;
XX
DR WPI; 1998-437388/37.
XX
PT Disease specific immunogen - comprises disease specific cytotoxic T
PT lymphocyte epitope used to elicit melanoma specific CTL response.
XX
PS Disclosure; Page 27; 93pp; English.
XX
CC The peptide epitope AAW77119-W77138 were created for human tumour-
CC specific cytotoxic T lymphocyte response. These peptides are are cysteine
CC - depleted mutants of a native disease-specific CTL epitope. The cysteine
CC - depleted CTL epitopes elicit a stronger or more specific CTL response
CC than the native epitope. The epitopes can be used in a disease-specific
CC immunogen to protect a mammal against disease in particular melanomas.
CC The peptides may also be used to screen a sample for the presence of an
CC antigen with the same epitope, or with a different cross-reactive epitope
XX
SQ Sequence 9 AA;

```
  Query Match             88.9%;  Score 40;  DB 2;  Length 9;
  Best Local Similarity   88.9%;  Pred. No. 1.8e+06;
  Matches    8;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 YLSGADLNL 9
              |||||:|||
Db          1 YLSGANLNL 9
```

Search completed: May 17, 2005, 06:17:49
Job time : 69 secs

GenCore version 5.1.6

OM protein - nucleic search, using frame_plus_p2n model

Run on: May 17, 2005, 16:32:09 ; Search time 1297.5 Seconds
(without alignments)
336.106 Million cell updates/sec

Title: US-10-725-373-2
Perfect score: 45
Sequence: 1 YLSGADLNL 9

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/US10725373/runat_17052005_071020_16185/app_query.fasta_1.796
-DB=GenEmbl -QFMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10725373_@CGN_1_1_6148_@runat_17052005_071020_16185 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	45	100.0	27	6	AR560605	AR560605 Sequence
	2	45	100.0	27	6	BD131676	BD131676 Carcinoem
	3	45	100.0	2106	6	AX133657	AX133657 Sequence
	4	45	100.0	2106	6	AX192349	AX192349 Sequence
	5	45	100.0	2106	6	AX393888	AX393888 Sequence
	6	43	95.6	27	6	AR560606	AR560606 Sequence
	7	43	95.6	27	6	BD131677	BD131677 Carcinoem
c	8	42	93.3	196838	2	AC115196	AC115196 Rattus no
c	9	42	93.3	282895	2	AC096056	AC096056 Rattus no
	10	40	88.9	27	6	AR560604	AR560604 Sequence
	11	40	88.9	27	6	BD131675	BD131675 Carcinoem
c	12	40	88.9	80	6	A87744	A87744 Sequence 38
c	13	40	88.9	80	6	BD195831	BD195831 Method fo
	14	40	88.9	155	6	AX193221	AX193221 Sequence
	15	40	88.9	256	6	AX260775	AX260775 Sequence
c	16	40	88.9	407	6	AX260391	AX260391 Sequence
	17	40	88.9	409	6	AR273297	AR273297 Sequence
	18	40	88.9	409	6	AR273585	AR273585 Sequence
c	19	40	88.9	409	6	AR273719	AR273719 Sequence
	20	40	88.9	409	6	AR276878	AR276878 Sequence
	21	40	88.9	409	6	AR277166	AR277166 Sequence
c	22	40	88.9	409	6	AR277300	AR277300 Sequence
	23	40	88.9	409	6	AR407153	AR407153 Sequence
	24	40	88.9	409	6	AR407441	AR407441 Sequence
c	25	40	88.9	409	6	AR407575	AR407575 Sequence
	26	40	88.9	409	6	AR441003	AR441003 Sequence
	27	40	88.9	409	6	AR441291	AR441291 Sequence
c	28	40	88.9	409	6	AR441425	AR441425 Sequence
	29	40	88.9	409	6	AR543814	AR543814 Sequence
	30	40	88.9	409	6	AR544102	AR544102 Sequence
c	31	40	88.9	409	6	AR544236	AR544236 Sequence
c	32	40	88.9	409	6	AX260450	AX260450 Sequence
c	33	40	88.9	409	6	AX351472	AX351472 Sequence
	34	40	88.9	409	6	AX368330	AX368330 Sequence
	35	40	88.9	409	6	AX368618	AX368618 Sequence
c	36	40	88.9	409	6	AX368752	AX368752 Sequence
	37	40	88.9	409	6	AX396625	AX396625 Sequence
	38	40	88.9	409	6	AX397282	AX397282 Sequence
c	39	40	88.9	410	6	AX260455	AX260455 Sequence
	40	40	88.9	410	6	AX260459	AX260459 Sequence
	41	40	88.9	412	6	AR166812	AR166812 Sequence
	42	40	88.9	412	6	BD265007	BD265007 Compounds
	43	40	88.9	412	6	AR400993	AR400993 Sequence
	44	40	88.9	412	6	AX192438	AX192438 Sequence
	45	40	88.9	413	6	AX260652	AX260652 Sequence

ALIGNMENTS

RESULT 1
AR560605
LOCUS AR560605 27 bp DNA linear PAT 08-OCT-2004
DEFINITION Sequence 7 from patent US 6756038.
ACCESSION AR560605
VERSION AR560605.1 GI:53972926
KEYWORDS .
SOURCE Unknown.
ORGANISM Unknown.
Unclassified.
REFERENCE 1 (bases 1 to 27)
AUTHORS Schlom,J., Barzaga,E. and Zaremba,S.
TITLE Agonist and antagonist peptides of carcinoembryonic antigen (CEA)
JOURNAL Patent: US 6756038-A 7 29-JUN-2004;
FEATURES Location/Qualifiers
source 1. .27
/organism="unknown"
/mol_type="genomic DNA"
ORIGIN

Alignment Scores:
Pred. No.: 0.00769 Length: 27
Score: 45.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0

```
US-10-725-373-2 (1-9) x AR560605 (1-27)

Qy          1 TyrLeuSerGlyAlaAspLeuAsnLeu 9
              ||||||||||||||||||||||||||
Db          1 TACCTTTCGGGAGCGGACCTCAACCTC 27

RESULT 2
BD131676
LOCUS       BD131676                  27 bp    DNA     linear   PAT 18-SEP-2002
DEFINITION  Carcinoembryonic antigen (CEA) agonist and antagonist peptides.
ACCESSION   BD131676
VERSION     BD131676.1  GI:23226621
KEYWORDS    JP 2002500002-A/2.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 27)
  AUTHORS   Schlom,J., Barzaga,E. and Zaremba,S.
  TITLE     Carcinoembryonic antigen (CEA) agonist and antagonist peptides
  JOURNAL   Patent: JP 2002500002-A 2 08-JAN-2002;
            THE UNITED STATES OF AMERICA
COMMENT     OS   Homo sapiens (human)
            PN   JP 2002500002-A/2
            PD   08-JAN-2002
            PF   22-SEP-1998 JP 2000516030
            PR   10-OCT-1997 US    60/061589
            PI   JEFFREY SCHLOM,ELENE BARZAGA,SAM ZAREMBA
            PC   C12N15/09,A61K38/00,A61K45/00,A61K48/00,A61P35/00,A61P37/02,
            PC   A61P43/00,
            PC   C07K14/705,C07K16/28,C12N5/10,C12N15/00,A61K37/02,C12N5/00 CC
             Carcinoembryonic antigen (CEA) agonist and antagonist peptides FH
              Key             Location/Qualifiers
            FT   source          1. .27
            FT                   /organism='Homo sapiens (human)'.
FEATURES              Location/Qualifiers
     source           1. .27
                      /organism="Homo sapiens"
                      /mol_type="genomic DNA"
                      /db_xref="taxon:9606"
ORIGIN

Alignment Scores:
Pred. No.:              0.00769         Length:         27
Score:                  45.00           Matches:        9
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     6               Gaps:           0

US-10-725-373-2 (1-9) x BD131676 (1-27)

Qy          1 TyrLeuSerGlyAlaAspLeuAsnLeu 9
              ||||||||||||||||||||||||||
Db          1 TACCTTTCGGGAGCGGACCTCAACCTC 27

RESULT 3
AX133657
LOCUS       AX133657                2106 bp    DNA     linear   PAT 15-MAY-2001
DEFINITION  Sequence 111 from Patent WO0130382.
ACCESSION   AX133657
VERSION     AX133657.1  GI:14139699
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1
  AUTHORS   Berinstein,N., Tartaglia,J., Moingeon,P. and Barber,B.
  TITLE     Method of inducing and/or enhancing an immune response to tumor
            antigens
  JOURNAL   Patent: WO 0130382-A 111 03-MAY-2001;
            Aventis Pasteur Limited (CA)
FEATURES              Location/Qualifiers
     source           1. .2106
                      /organism="synthetic construct"
                      /mol_type="unassigned DNA"
                      /db_xref="taxon:32630"
     CDS              1. .2106
                      /note="unnamed protein product; modified CEA"
                      /codon_start=1
                      /transl_table=11
                      /protein_id="CAC38955.1"
                      /db_xref="GI:14139700"
                      /translation="MESPSAPPHRWCIPWQRLLLTASLLTFWNPPTTAKLTIESTPFN
                      VAEGKEVLLLVHNLPQHLFGYSWYKGERVDGNRQIIGYVIGTQQATPGPAYSGREIIY
                      PNASLLIQNIIQNDTGFYTLHVIKSDLVNEEATGQFRVYPELPKPSISSNNSKPVEDK
                      DAVAFTCEPETQDATYLWWVNNQSLPVSPRLQLSNGNRTLTLFNVTRNDTASYKCETQ
                      NPVSARRSDSVILNVLYGPDAPTISPLNTSYRSGENLNLSCHAASNPPAQYSWFVNGT
                      FQQSTQELFIPNITVNNSGSYTCQAHNSDTGLNRTTVTTITVYEPPKPFITSNNSNPV
                      EDEDAVALTCEPEIQNTTYLWWVNNQSLPVSPRLQLSNDNRTLTLLSVTRNDVGPYEC
                      GIQNELSVDHSDPVILNVLYGPDDPTISPSYTYYRPGVNLSLSCHAASNPPAQYSWLI
                      DGNIQQHTQELFISNITEKNSGLYTCQANNSASGHSRTTVKTITVSAELPKPSISSNN
                      SKPVEDKDAVAFTCEPEAQNTTYLWWVNGQSLPVSPRLQLSNGNRTLTLFNVTRNDAR
                      AYVCGIQNSVSANRSDPVTLDVLYGPDTPIISPPDSSYLSGADLNLSCHSASNPSPQY
                      SWRINGIPQQHTQVLFIAKITPNNNGTYACFVSNLATGRNNSIVKSITVSASGTSPGL
                      SAGATVGIMIGVLVGVALI"
ORIGIN

Alignment Scores:
Pred. No.:              1.25            Length:         2106
Score:                  45.00           Matches:        9
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     6               Gaps:           0

US-10-725-373-2 (1-9) x AX133657 (1-2106)

Qy          1 TyrLeuSerGlyAlaAspLeuAsnLeu 9
              ||||||||||||||||||||||||||
Db       1810 TACCTTTCGGGAGCGGACCTCAACCTC 1836

RESULT 4
AX192349
LOCUS       AX192349                2106 bp    DNA     linear   PAT 15-AUG-2001
DEFINITION  Sequence 3 from Patent WO0149317.
ACCESSION   AX192349
VERSION     AX192349.1  GI:15210326
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1
  AUTHORS   Emtage,P., Barber,B.H., Sambhara,S. and Sia,C.D.
  TITLE     Enhancing the immune response to an antigen by presensitizing with
            an inducing agent prior to immunizing with the inducing agent and
            the antigen
  JOURNAL   Patent: WO 0149317-A 3 12-JUL-2001;
            Aventis Pasteur Limited (CA)
FEATURES              Location/Qualifiers
     source           1. .2106
                      /organism="synthetic construct"
                      /mol_type="unassigned DNA"
                      /db_xref="taxon:32630"
     CDS              1. .2106
                      /note="unnamed protein product; modified CEA"
                      /codon_start=1
                      /transl_table=11
                      /protein_id="CAC51319.1"
                      /db_xref="GI:15210327"
                      /translation="MESPSAPPHRWCIPWQRLLLTASLLTFWNPPTTAKLTIESTPFN
                      VAEGKEVLLLVHNLPQHLFGYSWYKGERVDGNRQIIGYVIGTQQATPGPAYSGREIIY
                      PNASLLIQNIIQNDTGFYTLHVIKSDLVNEEATGQFRVYPELPKPSISSNNSKPVEDK
                      DAVAFTCEPETQDATYLWWVNNQSLPVSPRLQLSNGNRTLTLFNVTRNDTASYKCETQ
                      NPVSARRSDSVILNVLYGPDAPTISPLNTSYRSGENLNLSCHAASNPPAQYSWFVNGT
                      FQQSTQELFIPNITVNNSGSYTCQAHNSDTGLNRTTVTTITVYEPPKPFITSNNSNPV
```

EDEDAVALTCEPEIQNTTYLWWVNNQSLPVSPRLQLSNDNRTLTLLSVTRNDVGPYEC
GIQNELSVDHSDPVILNVLYGPDDPTISPSYTYYRPGVNLSLSCHAASNPPAQYSWLI
DGNIQQHTQELFISNITEKNSGLYTCQANNSASGHSRTTVKTITVSAELPKPSISSNN
SKPVEDKDAVAFTCEPEAQNTTYLWWVNGQSLPVSPRLQLSNGNRTLTLFNVTRNDAR
AYVCGIQNSVSANRSDPVTLDVLYGPDTPIISPPDSSYLSGADLNLSCHSASNPSPQY
SWRINGIPQQHTQVLFIAKITPNNNGTYACFVSNLATGRNNSIVKSITVSASGTSPGL
SAGATVGIMIGVLVGVALI"

ORIGIN

Alignment Scores:

Pred. No.:	1.25	Length:	2106
Score:	45.00	Matches:	9
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	6	Gaps:	0

US-10-725-373-2 (1-9) x AX192349 (1-2106)

```
Qy          1 TyrLeuSerGlyAlaAspLeuAsnLeu 9
              |||||||||||||||||||||||||||
Db       1810 TACCTTTCGGGAGCGGACCTCAACCTC 1836
```

RESULT 5
AX393888
LOCUS AX393888 2106 bp DNA linear PAT 23-MAR-2002
DEFINITION Sequence 2 from Patent WO0210379.
ACCESSION AX393888
VERSION AX393888.1 GI:19701852
KEYWORDS .
SOURCE synthetic construct
ORGANISM synthetic construct
other sequences; artificial sequences.
REFERENCE 1
AUTHORS Berinstein,N., Tartaglia,J., Tine,J.A., Panicali,D.L., Gritz,L. and Schlom,J.
TITLE Modified cea and uses thereof
JOURNAL Patent: WO 0210379-A 2 07-FEB-2002;
Aventis Pasteur Limited (CA) ; Therion Biologics (US) ; National Cancer Institute (US)
FEATURES Location/Qualifiers
source 1. .2106
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/note="Cea modified polypeptide"
ORIGIN

Alignment Scores:

Pred. No.:	1.25	Length:	2106
Score:	45.00	Matches:	9
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	6	Gaps:	0

US-10-725-373-2 (1-9) x AX393888 (1-2106)

```
Qy          1 TyrLeuSerGlyAlaAspLeuAsnLeu 9
              |||||||||||||||||||||||||||
Db       1810 TACCTTTCGGGAGCGGACCTCAACCTC 1836
```

RESULT 6
AR560606
LOCUS AR560606 27 bp DNA linear PAT 08-OCT-2004
DEFINITION Sequence 8 from patent US 6756038.
ACCESSION AR560606
VERSION AR560606.1 GI:53972927
KEYWORDS .
SOURCE Unknown.
ORGANISM Unknown.
Unclassified.
REFERENCE 1 (bases 1 to 27)
AUTHORS Schlom,J., Barzaga,E. and Zaremba,S.
TITLE Agonist and antagonist peptides of carcinoembryonic antigen (CEA)
JOURNAL Patent: US 6756038-A 8 29-JUN-2004;
FEATURES Location/Qualifiers
source 1. .27
/organism="unknown"
/mol_type="genomic DNA"
ORIGIN

Alignment Scores:

Pred. No.:	0.0244	Length:	27
Score:	43.00	Matches:	8
Percent Similarity:	100.00%	Conservative:	1
Best Local Similarity:	88.89%	Mismatches:	0
Query Match:	95.56%	Indels:	0
DB:	6	Gaps:	0

US-10-725-373-2 (1-9) x AR560606 (1-27)

```
Qy          1 TyrLeuSerGlyAlaAspLeuAsnLeu 9
              |||||||||||||||||:::|||||||
Db          1 TACCTTTCGGGAGCGGACATCAACCTC 27
```

RESULT 7
BD131677
LOCUS BD131677 27 bp DNA linear PAT 18-SEP-2002
DEFINITION Carcinoembryonic antigen (CEA) agonist and antagonist peptides.
ACCESSION BD131677
VERSION BD131677.1 GI:23226622
KEYWORDS JP 2002500002-A/3.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 27)
AUTHORS Schlom,J., Barzaga,E. and Zaremba,S.
TITLE Carcinoembryonic antigen (CEA) agonist and antagonist peptides
JOURNAL Patent: JP 2002500002-A 3 08-JAN-2002;
THE UNITED STATES OF AMERICA
COMMENT OS Homo sapiens (human)
PN JP 2002500002-A/3
PD 08-JAN-2002
PF 22-SEP-1998 JP 2000516030
PR 10-OCT-1997 US 60/061589
PI JEFFREY SCHLOM,ELENE BARZAGA,SAM ZAREMBA
PC C12N15/09,A61K38/00,A61K45/00,A61K48/00,A61P35/00,A61P37/02,
PC A61P43/00,
PC C07K14/705,C07K16/28,C12N5/10,C12N15/00,A61K37/02,C12N5/00 CC
Carcinoembryonic antigen (CEA) agonist and antagonist peptides FH
Key Location/Qualifiers
FT source 1. .27
FT /organism='Homo sapiens (human)'.
FEATURES Location/Qualifiers
source 1. .27
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
ORIGIN

Alignment Scores:

Pred. No.:	0.0244	Length:	27
Score:	43.00	Matches:	8
Percent Similarity:	100.00%	Conservative:	1
Best Local Similarity:	88.89%	Mismatches:	0
Query Match:	95.56%	Indels:	0
DB:	6	Gaps:	0

US-10-725-373-2 (1-9) x BD131677 (1-27)

```
Qy          1 TyrLeuSerGlyAlaAspLeuAsnLeu 9
              |||||||||||||||||:::|||||||
Db          1 TACCTTTCGGGAGCGGACATCAACCTC 27
```

```
RESULT 8
AC115196/c
LOCUS       AC115196               196838 bp    DNA     linear   HTG 19-NOV-2002
DEFINITION  Rattus norvegicus clone CH230-291E8, WORKING DRAFT SEQUENCE, 4
            unordered pieces.
ACCESSION   AC115196
VERSION     AC115196.4  GI:25072711
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 196838)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 196838)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-MAR-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 196838)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-NOV-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Nov 19, 2002 this sequence version replaced gi:23618799.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GSYI
                Center clone name: CH230-291E8
            -------------- Summary Statistics
                Assembly program: Phrap; version 0.990329
                Consensus quality: 175442 bases at least Q40
                Consensus quality: 177660 bases at least Q30
                Consensus quality: 179434 bases at least Q20
                Estimated insert size: 177868; sum-of-contigs estimation
                Quality coverage: 9x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 4 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    39826: contig of 39826 bp in length
            *    39827    39926: gap of unknown length
            *    39927   194015: contig of 154089 bp in length
            *   194016   194115: gap of unknown length
            *   194116   195420: contig of 1305 bp in length
            *   195421   195520: gap of unknown length
            *   195521   196838: contig of 1318 bp in length.
FEATURES             Location/Qualifiers
     source          1. .196838
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-291E8"
     misc_feature    complement(32622. .33371)
                     /note="clone_boundary
                     clone_end:Sp6
                     site:
                     end_sequence:BZ166109"
     misc_feature    35496. .36601
                     /note="wgs_contig"
     misc_feature    38810. .39826
                     /note="wgs_contig"
     misc_feature    39927. .43082
                     /note="wgs_contig"
ORIGIN

Alignment Scores:
Pred. No.:              1.42e+03        Length:         196838
Score:                  42.00           Matches:        8
Percent Similarity:     100.00%         Conservative:   1
```

```
Best Local Similarity:  88.89%        Mismatches:     0
Query Match:            93.33%        Indels:         0
DB:                     2             Gaps:           0

US-10-725-373-2 (1-9) x AC115196 (1-196838)

Qy          1 TyrLeuSerGlyAlaAspLeuAsnLeu 9
              ||||||||||||:::|||||||||||
Db     162369 TACCTCTCAGGTTCAGACTTAAACCTA 162343

RESULT 9
AC096056/c
LOCUS       AC096056               282895 bp    DNA     linear   HTG 09-NOV-2002
DEFINITION  Rattus norvegicus clone CH230-22P2, WORKING DRAFT SEQUENCE, 2
            unordered pieces.
ACCESSION   AC096056
VERSION     AC096056.5  GI:24817935
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 282895)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 282895)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-SEP-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 282895)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-NOV-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Nov 9, 2002 this sequence version replaced gi:23322196.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GEIC
                Center clone name: CH230-22P2
            -------------- Summary Statistics
                Assembly program: Phrap; version 0.990329
                Consensus quality: 237332 bases at least Q40
                Consensus quality: 239717 bases at least Q30
                Consensus quality: 241152 bases at least Q20
                Estimated insert size: 249163; sum-of-contigs estimation
                Quality coverage: 7x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 2 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *         1   246103: contig of 246103 bp in length
            *    246104   246203: gap of unknown length
            *    246204   282895: contig of 36692 bp in length.
FEATURES             Location/Qualifiers
     source          1. .282895
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-22P2"
     misc_feature    1. .1804
                     /note="wgs_end_extension
                     clone_end:T7"
     misc_feature    complement(2727. .3647)
                     /note="clone_boundary
                     clone_end:T7
                     site:EcoRI
                     end_sequence:BH360868"
     misc_feature    240809. .240940
                     /note="clone_boundary
                     clone_end:Sp6
```

```
                     site:EcoRI
                     end_sequence:BH360870"
     misc_feature    244238. .246103
                     /note="wgs_end_extension
                     clone_end:Sp6"
ORIGIN

Alignment Scores:
Pred. No.:                2.17e+03       Length:          282895
Score:                    42.00          Matches:         8
Percent Similarity:       100.00%        Conservative:    1
Best Local Similarity:    88.89%         Mismatches:      0
Query Match:              93.33%         Indels:          0
DB:                       2              Gaps:            0

US-10-725-373-2 (1-9) x AC096056 (1-282895)

Qy          1 TyrLeuSerGlyAlaAspLeuAsnLeu 9
              ||||||||||||:::|||||||||||||
Db      20770 TACCTCTCAGGTTCAGACTTAAACCTA 20744

RESULT 10
AR560604
LOCUS       AR560604                  27 bp    DNA     linear   PAT 08-OCT-2004
DEFINITION  Sequence 6 from patent US 6756038.
ACCESSION   AR560604
VERSION     AR560604.1  GI:53972925
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 27)
  AUTHORS   Schlom,J., Barzaga,E. and Zaremba,S.
  TITLE     Agonist and antagonist peptides of carcinoembryonic antigen (CEA)
  JOURNAL   Patent: US 6756038-A 6 29-JUN-2004;
FEATURES             Location/Qualifiers
     source          1. .27
                     /organism="unknown"
                     /mol_type="genomic DNA"
ORIGIN

Alignment Scores:
Pred. No.:                0.138          Length:          27
Score:                    40.00          Matches:         8
Percent Similarity:       100.00%        Conservative:    1
Best Local Similarity:    88.89%         Mismatches:      0
Query Match:              88.89%         Indels:          0
DB:                       6              Gaps:            0

US-10-725-373-2 (1-9) x AR560604 (1-27)

Qy          1 TyrLeuSerGlyAlaAspLeuAsnLeu 9
              ||||||||||||||:::|||||||||
Db          1 TACCTTTCGGGAGCGAACCTCAACCTC 27

RESULT 11
BD131675
LOCUS       BD131675                  27 bp    DNA     linear   PAT 18-SEP-2002
DEFINITION  Carcinoembryonic antigen (CEA) agonist and antagonist peptides.
ACCESSION   BD131675
VERSION     BD131675.1  GI:23226620
KEYWORDS    JP 2002500002-A/1.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 27)
  AUTHORS   Schlom,J., Barzaga,E. and Zaremba,S.
  TITLE     Carcinoembryonic antigen (CEA) agonist and antagonist peptides
  JOURNAL   Patent: JP 2002500002-A 1 08-JAN-2002;
            THE UNITED STATES OF AMERICA
COMMENT     OS   Homo sapiens (human)
            PN   JP 2002500002-A/1
            PD   08-JAN-2002
            PF   22-SEP-1998 JP 2000516030
            PR   10-OCT-1997 US    60/061589
            PI   JEFFREY SCHLOM,ELENE BARZAGA,SAM ZAREMBA
            PC   C12N15/09,A61K38/00,A61K45/00,A61K48/00,A61P35/00,A61P37/02,
            PC   A61P43/00,
            PC   C07K14/705,C07K16/28,C12N5/10,C12N15/00,A61K37/02,C12N5/00 CC
             Carcinoembryonic antigen (CEA) agonist and antagonist peptides FH
             Key             Location/Qualifiers
            FT   source          1. .27
            FT                   /organism='Homo sapiens (human)'.
FEATURES             Location/Qualifiers
     source          1. .27
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
ORIGIN

Alignment Scores:
Pred. No.:                0.138          Length:          27
Score:                    40.00          Matches:         8
Percent Similarity:       100.00%        Conservative:    1
Best Local Similarity:    88.89%         Mismatches:      0
Query Match:              88.89%         Indels:          0
DB:                       6              Gaps:            0

US-10-725-373-2 (1-9) x BD131675 (1-27)

Qy          1 TyrLeuSerGlyAlaAspLeuAsnLeu 9
              ||||||||||||||:::|||||||||
Db          1 TACCTTTCGGGAGCGAACCTCAACCTC 27

RESULT 12
A87744/c
LOCUS       A87744                    80 bp    DNA     linear   PAT 22-JAN-2000
DEFINITION  Sequence 38 from Patent WO9833523.
ACCESSION   A87744
VERSION     A87744.1  GI:6736346
KEYWORDS    .
SOURCE      unidentified
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 80)
  AUTHORS   Carr,F.J. and Carter,G.
  TITLE     VACCINATION METHODS AND MOLECULES
  JOURNAL   Patent: WO 9833523-A 38 06-AUG-1998;
            BIOVATION LIMITED (GB); CARR FRANK JOSEPH (GB)
FEATURES             Location/Qualifiers
     source          1. .80
                     /organism="unidentified"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32644"
ORIGIN

Alignment Scores:
Pred. No.:                0.49           Length:          80
Score:                    40.00          Matches:         8
Percent Similarity:       100.00%        Conservative:    1
Best Local Similarity:    88.89%         Mismatches:      0
Query Match:              88.89%         Indels:          0
DB:                       6              Gaps:            0

US-10-725-373-2 (1-9) x A87744 (1-80)

Qy          1 TyrLeuSerGlyAlaAspLeuAsnLeu 9
              ||||||||||||||:::|||||||||
Db         68 TACCTGTCCGGCGCCAACCTGAACCTG 42

RESULT 13
BD195831/c
LOCUS       BD195831                  80 bp    DNA     linear   PAT 17-JUL-2003
DEFINITION  Method for the production of non-immunogenic proteins.
ACCESSION   BD195831
```

```
VERSION     BD195831.1  GI:33005601
KEYWORDS    JP 2002512624-A/102.
SOURCE      unidentified
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 80)
  AUTHORS   Carr,F.J., Adair,F.S., Hamilton,A.A. and Carter,G.
  TITLE     Method for the production of non-immunogenic proteins
  JOURNAL   Patent: JP 2002512624-A 102 23-APR-2002;
            BIOVATION LTD
COMMENT     OS   Unidentified
            PN   JP 2002512624-A/102
            PD   23-APR-2002
            PF   21-MAY-1998 JP 1998550129
            PR   21-MAY-1997 GB    9710480.6,31-JUL-1997 GB    9716197.0 PR
            28-NOV-1997 GB    9725270.4,02-DEC-1997 US    60/067235 PR
            14-APR-1998 GB    9807751.4
            PI   FRANCIS JOSEPH CARR,FIONA SUZANNE ADAIR,ANITA ANNE HAMILTON,
            PI   GRAHAM CARTER
            PC   C07K16/46,C07K14/315,G01N33/563,A61K39/395
            CC   Strandedness: Single;
            CC   Topology: Linear;
            CC   Method for the production of non-immunogenic proteins FH   Key
                       Location/Qualifiers
            FT   source          1. .80
            FT                   /organism='Unidentified'.
FEATURES              Location/Qualifiers
     source           1. .80
                      /organism="unidentified"
                      /mol_type="genomic DNA"
                      /db_xref="taxon:32644"
ORIGIN

Alignment Scores:
Pred. No.:              0.49          Length:         80
Score:                  40.00         Matches:        8
Percent Similarity:     100.00%       Conservative:   1
Best Local Similarity:  88.89%        Mismatches:     0
Query Match:            88.89%        Indels:         0
DB:                     6             Gaps:           0

US-10-725-373-2 (1-9) x BD195831 (1-80)

Qy          1 TyrLeuSerGlyAlaAspLeuAsnLeu 9
              |||||||||||||||:::|||||||||
Db         68 TACCTGTCCGGCGCCAACCTGAACCTG 42

RESULT 14
AX193221
LOCUS       AX193221                155 bp    DNA     linear   PAT 15-AUG-2001
DEFINITION  Sequence 788 from Patent WO0149716.
ACCESSION   AX193221
VERSION     AX193221.1  GI:15211172
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Xu,J., Lodes,M.J., Secrist,H., Benson,D.R., Meagher,M.J.,
            Stolk,J.A., King,G.E., Wang,T. and Jiang,Y.
  TITLE     Compounds for immunotherapy and diagnosis of colon cancer and
            methods for their use
  JOURNAL   Patent: WO 0149716-A 788 12-JUL-2001;
            CORIXA CORPORATION (US)
FEATURES              Location/Qualifiers
     source           1. .155
                      /organism="Homo sapiens"
                      /mol_type="unassigned DNA"
                      /db_xref="taxon:9606"
ORIGIN

Alignment Scores:
Pred. No.:              1.06          Length:         155
Score:                  40.00         Matches:        8
Percent Similarity:     100.00%       Conservative:   1
Best Local Similarity:  88.89%        Mismatches:     0
Query Match:            88.89%        Indels:         0
DB:                     6             Gaps:           0

US-10-725-373-2 (1-9) x AX193221 (1-155)

Qy          1 TyrLeuSerGlyAlaAspLeuAsnLeu 9
              |||||||||||||||:::|||||||||
Db        119 TACCTTTCNGGAGCGAACCTCAACCTC 145

RESULT 15
AX260775
LOCUS       AX260775                256 bp    DNA     linear   PAT 26-OCT-2001
DEFINITION  Sequence 426 from Patent WO0173027.
ACCESSION   AX260775
VERSION     AX260775.1  GI:16509742
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Meagher,M.J., Xu,J. and King,G.E.
  TITLE     Compositions and methods for therapy and diagnosis of colon cancer
  JOURNAL   Patent: WO 0173027-A 426 04-OCT-2001;
            CORIXA CORPORATION (US)
FEATURES              Location/Qualifiers
     source           1. .256
                      /organism="Homo sapiens"
                      /mol_type="unassigned DNA"
                      /db_xref="taxon:9606"
ORIGIN

Alignment Scores:
Pred. No.:              1.91          Length:         256
Score:                  40.00         Matches:        8
Percent Similarity:     100.00%       Conservative:   1
Best Local Similarity:  88.89%        Mismatches:     0
Query Match:            88.89%        Indels:         0
DB:                     6             Gaps:           0

US-10-725-373-2 (1-9) x AX260775 (1-256)

Qy          1 TyrLeuSerGlyAlaAspLeuAsnLeu 9
              |||||||||||||||:::|||||||||
Db         15 TACCTTTCGGGAGCGAACCTCAACCTC 41


Search completed: May 17, 2005, 19:12:10
Job time : 1327.5 secs
```

GenCore version 5.1.6

OM protein - nucleic search, using frame_plus_p2n model

Run on: May 17, 2005, 16:29:39 ; Search time 326.5 Seconds
(without alignments)
163.178 Million cell updates/sec

Title: US-10-725-373-2
Perfect score: 45
Sequence: 1 YLSGADLNL 9

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/US10725373/runat_17052005_071020_16177/app_query.fasta_1.796
-DB=N_Geneseq_16Dec04 -QFMT=fastap -SUFFIX=rng -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10725373_@CGN_1_1_1241_@runat_17052005_071020_16177 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N_Geneseq_16Dec04:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002as:*
7: geneseqn2002bs:*
8: geneseqn2003as:*
9: geneseqn2003bs:*
10: geneseqn2003cs:*
11: geneseqn2003ds:*
12: geneseqn2004as:*
13: geneseqn2004bs:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	45	100.0	27	2	AAX56258	Aax56258 Carcinoem
2	45	100.0	2105	6	AAI72497	Aai72497 CEA agoni
3	45	100.0	2106	4	AAH20121	Aah20121 Modified
4	45	100.0	2106	5	AAD07347	Aad07347 Modified
5	45	100.0	2106	6	AAI72489	Aai72489 CEA agoni
6	45	100.0	2106	10	ADE13860	Ade13860 CEA-CAP6D
7	45	100.0	2106	10	ADE13861	Ade13861 CEA(6D)-1
c 8	45	100.0	7958	6	AAI72490	Aai72490 H6-promot
9	43	95.6	27	2	AAX56259	Aax56259 Carcinoem
10	40	88.9	27	2	AAX56260	Aax56260 Carcinoem
11	40	88.9	30	12	ADL46174	Adl46174 Human CAP
12	40	88.9	64	12	ADL46175	Adl46175 Human imm
c 13	40	88.9	80	2	AAV57948	Aav57948 708 vkcea
c 14	40	88.9	80	2	AAV81101	Aav81101 Vaccine 2
15	40	88.9	155	4	AAI29234	Aai29234 Colon tum
16	40	88.9	155	8	ABZ33420	Abz33420 Human col
17	40	88.9	256	4	AAS57750	Aas57750 cDNA #426
c 18	40	88.9	340	6	ABV88334	Abv88334 Human col
c 19	40	88.9	407	4	AAS57366	Aas57366 cDNA #42
c 20	40	88.9	409	4	AAS57425	Aas57425 cDNA #101
21	40	88.9	409	6	ABV86774	Abv86774 Human col
c 22	40	88.9	409	6	ABV87551	Abv87551 Human col
23	40	88.9	409	6	ABV89100	Abv89100 Human col
24	40	88.9	409	6	ABV87855	Abv87855 Human col
25	40	88.9	409	6	ABK39290	Abk39290 DNA encod
26	40	88.9	409	6	ABK39002	Abk39002 cDNA enco
c 27	40	88.9	409	6	ABK39424	Abk39424 DNA encod
28	40	88.9	409	6	ABK45946	Abk45946 cDNA enco
29	40	88.9	409	6	ABK45289	Abk45289 cDNA enco
c 30	40	88.9	409	6	ABK27782	Abk27782 Human col
31	40	88.9	409	8	ACA11619	Aca11619 Human lun
32	40	88.9	409	8	ACA11331	Aca11331 Human lun
c 33	40	88.9	409	8	ACA11753	Aca11753 Human lun
c 34	40	88.9	409	8	ACA02939	Aca02939 Lung canc
35	40	88.9	409	8	ACA02805	Aca02805 Lung canc
36	40	88.9	409	8	ACA02517	Aca02517 Lung canc
37	40	88.9	409	10	ADH46559	Adh46559 Human lun
38	40	88.9	409	10	ADH46847	Adh46847 Human lun
c 39	40	88.9	409	10	ADH46981	Adh46981 Human lun
40	40	88.9	409	13	ADJ20766	Adj20766 Human lun
41	40	88.9	409	13	ADJ20478	Adj20478 Human lun
c 42	40	88.9	409	13	ADJ20900	Adj20900 Human lun
c 43	40	88.9	410	4	AAS57430	Aas57430 cDNA #106
44	40	88.9	410	4	AAS57434	Aas57434 cDNA #110
c 45	40	88.9	410	6	ABV88264	Abv88264 Human col

ALIGNMENTS

RESULT 1
AAX56258
ID AAX56258 standard; DNA; 27 BP.
XX
AC AAX56258;
XX
DT 20-JUL-1999 (first entry)
XX
DE Carcinoembryonic antigen peptide agonist encoding DNA SEQ ID NO:7.
XX
KW Carcinoembryonic antigen; CEA; human; agonist; antagonist;
KW immune response; carcinoma; gastrointestinal; breast; pancreatic;
KW bladder; ovarian; lung; prostatic; T cell proliferation; cancer;
KW adoptive transfer therapy; autoimmune reaction; immunotherapy; ss.
XX
OS Homo sapiens.
OS Synthetic.
XX
PN WO9919478-A1.
XX
PD 22-APR-1999.
XX
PF 22-SEP-1998; 98WO-US019794.
XX
PR 10-OCT-1997; 97US-0061589P.
XX
PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
XX
PI Schlom J, Barzaga E, Zaremba S;

XX
DR WPI; 1999-326544/27.
XX
PT Peptide agonists and antagonists of carcinoembryonal antigen.
XX
PS Claim 22; Page 20; 72pp; English.
XX
CC The present invention describes peptides (A) that comprise agonists (Ia)
CC or antagonists (Ib) of human carcinoembryonal antigen (CEA). (Ia) are
CC used in vaccines to kill or inhibit carcinoma cells that express CEA or
CC its epitopes, particularly for treating gastrointestinal, breast,
CC pancreatic, bladder, ovarian, lung or prostatic carcinoma. They can also
CC be used to proliferate T cells, e.g. from vaccinated subjects, for use in
CC adoptive transfer therapy. (Ib) are used to inhibit CEA-specific immune
CC responses, e.g. in vaccinated subjects, to prevent an autoimmune reaction
CC to cancer immunotherapy (i.e. to prevent attack on normal but CEA-
CC expressing cells). (Ia) are more active than native sequence (I) and
CC generate a highly specific and systemic anti-CEA response. Cytotoxic T
CC cells generated recognize both (Ia) and native CEA epitopes. The present
CC sequence encodes a specifically claimed example of (Ia)
XX
SQ Sequence 27 BP; 5 A; 10 C; 6 G; 6 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	0.0518	Length:	27
Score:	45.00	Matches:	9
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	2	Gaps:	0

US-10-725-373-2 (1-9) x AAX56258 (1-27)

```
Qy          1 TyrLeuSerGlyAlaAspLeuAsnLeu 9
              |||||||||||||||||||||||||||
Db          1 TACCTTTCGGGAGCGGACCTCAACCTC 27
```

RESULT 2
AAI72497
ID AAI72497 standard; DNA; 2105 BP.
XX
AC AAI72497;
XX
DT 16-MAY-2002 (first entry)
XX
DE CEA agonist coding sequence #2.
XX
KW CAP-1; epitope; carcinoembryonic antigen; CEA; agonist; immune response;
KW carcinoma; gastrointestinal; breast; pancreatic; bladder; ovarian; lung;
KW prostate; cancer; gene; ss.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT CDS 1. .2105
FT /*tag= a
FT /product= "CEA agonist polypeptide"
XX
PN WO200210379-A2.
XX
PD 07-FEB-2002.
XX
PF 27-JUL-2001; 2001WO-CA001092.
XX
PR 31-JUL-2000; 2000US-0222043P.
XX
PA (AVET) AVENTIS PASTEUR LTD.
PA (THER-) THERION BIOLOGICS.
PA (USSH) US NAT CANCER INST.
XX
PI Berinstein N, Tartaglia J, Tine JA, Panicali DL, Gritz L;
PI Schlom J;
XX
DR WPI; 2002-206189/26.
XX
PT Carcinoembryonic antigen agonist polypeptide for inducing an immune
PT response in animal against antigen and for inhibiting an epitope antigen
PT expressing carcinoma cell, comprises a modified antigen epitope.
XX
PS Claim 4; Page 66-67; 69pp; English.
XX
CC This sequence encodes the carcinoembryonic antigen (CEA) agonist
CC polypeptide of the invention. This sequence represents the sequence given
CC in the Seq ID listing in the specification, and does not directly encodes
CC the CEA agonist polypeptide given in AAB47919. The CEA agonist contains a
CC modified CAP-1 epitope of CEA, in which position 6 is modified from Asp
CC to Asn to increase its immunogenicity. The CEA agonist polypeptide of the
CC invention, or DNA encoding it, are useful for: (i) inducing an immune
CC response in an animal directed against a CEA protein or fragment, CEA
CC agonist, a CEA epitope, a modified CEA epitope, cells expressing or
CC binding a CEA protein or fragment; and (ii) inhibiting a CEA epitope
CC expressing carcinoma cell, which is a gastrointestinal, breast,
CC pancreatic, bladder, ovarian, lung or prostate carcinoma cell in a
CC patient, hence is useful for manufacture of a medicament for the
CC treatment of cancer
XX
SQ Sequence 2105 BP; 555 A; 658 C; 441 G; 451 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	7.05	Length:	2105
Score:	45.00	Matches:	9
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	6	Gaps:	0

US-10-725-373-2 (1-9) x AAI72497 (1-2105)

```
Qy          1 TyrLeuSerGlyAlaAspLeuAsnLeu 9
              |||||||||||||||||||||||||||
Db       1809 TACCTTTCGGGAGCGGACCTCAACCTC 1835
```

RESULT 3
AAH20121
ID AAH20121 standard; cDNA; 2106 BP.
XX
AC AAH20121;
XX
DT 08-AUG-2001 (first entry)
XX
DE Modified Carcinoembryonic antigen (CEA) encoding cDNA SEQ ID NO:111.
XX
KW Virus; adenovirus; poxvirus; alphavirus; immune response; gp100;
KW tumour antigen; CEA; carcinoembryonic antigen; immunostimulant;
KW cytostatic; immunotherapy; interferon-gamma; IFN-gamma; cancer; ss.
XX
OS Unidentified.
XX
PN WO200130382-A1.
XX
PD 03-MAY-2001.
XX
PF 20-OCT-2000; 2000WO-CA001253.
XX
PR 22-OCT-1999; 99US-0160879P.
PR 07-AUG-2000; 2000US-0223325P.
XX
PA (AVET) AVENTIS PASTEUR LTD.
XX
PI Berinstein N, Tartaglia J, Moingeon P, Barber B;
XX
DR WPI; 2001-308587/32.
DR P-PSDB; AAB97817.
XX
PT Inducing immune response to tumor antigen, useful in immunotherapy of
PT cancer, by administering the antigen to a lymphatic site.

```
XX
PS   Disclosure; Fig 8; 60pp; English.
XX
CC   The present invention describes a method for inducing an immune response,
CC   in an animal, to a tumour antigen (Ag) comprising administering Ag, or
CC   nucleic acid (I) that encodes it, to a lymphatic site. Cynomolgus monkeys
CC   (Macaca fascicularis) were injected with a modified form of gp100 antigen
CC   (a) into the left inguinal lymph node or (b) subcutaneously. Both animals
CC   of (a) developed a cell-mediated response (indicated by production of
CC   interferon-gamma from T lymphocytes when exposed to gp100 peptides), but
CC   only 2 of 4 animals of (b) did so. Also animals in (a) produced a far
CC   greater antibody response to gp100. The method is used in immunotherapy
CC   of a wide range of cancers through induction of a specific immune
CC   response (humoral and cellular) against the tumour antigens. When
CC   administered to a lymphatic site, Ag (or (I)) induces a stronger immune
CC   response than administration by other routes and may also break tolerance
CC   to Ag. AAB97708 and AAB97709 represent gp100 epitopes; AAB97710 to
CC   AAB97815 represent peptides derived from gp100 which stimulate interferon
CC   (IFN)-gamma production; AAH20120 encodes the modified gp100 protein given
CC   in AAB97816; AAH20121 encodes the modified carcinoembryonic antigen (CEA)
CC   protein given in AAB97817; and AAB97818 represents a CEA modified antigen
CC   peptide, all of which are used in the exemplification of the present
CC   invention
XX
SQ   Sequence 2106 BP; 559 A; 659 C; 442 G; 446 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:                7.06          Length:          2106
Score:                    45.00         Matches:         9
Percent Similarity:       100.00%       Conservative:    0
Best Local Similarity:    100.00%       Mismatches:      0
Query Match:              100.00%       Indels:          0
DB:                       4             Gaps:            0

US-10-725-373-2 (1-9) x AAH20121 (1-2106)

Qy          1 TyrLeuSerGlyAlaAspLeuAsnLeu 9
              |||||||||||||||||||||||||||
Db       1810 TACCTTTCGGGAGCGGACCTCAACCTC 1836

RESULT 4
AAD07347
ID   AAD07347 standard; DNA; 2106 BP.
XX
AC   AAD07347;
XX
DT   18-SEP-2001  (first entry)
XX
DE   Modified carcinoembryonic antigen (CEA) DNA.
XX
KW   Tumour-associated antigen; TAA; cytostatic; vaccine; gene therapy;
KW   immune response; tetanus toxoid; TT; diphtheria toxoid; DT; prophylactic;
KW   cancer; therapeutic; carcinoembryonic antigen; CEA; ds.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   CDS             1. .2106
FT                   /*tag=  a
FT                   /product= "Modified carcinoembryonic antigen (CEA)"
XX
PN   WO200149317-A2.
XX
PD   12-JUL-2001.
XX
PF   05-JAN-2001; 2001WO-CA000005.
XX
PR   05-JAN-2000; 2000US-0174587P.
XX
PA   (AVET ) AVENTIS PASTEUR LTD.
XX
PI   Emtage P,  Barber BH,  Sambhara S,  Sia CDY;
XX
DR   WPI; 2001-441790/47.
DR   P-PSDB; AAE05117.
XX
PT   Enhancing immune response to antigen such as tumor antigen for treating
PT   cancer in an animal involves administering an inducing agent to the
PT   animal followed by administering inducing agent-antigen mixture.
XX
PS   Claim 9; Fig 3; 62pp; English.
XX
CC   The invention relates to a method of enhancing an immune response against
CC   tumour-associated antigens (TAAs), such as GP100 and carcinoembryonic
CC   antigen (CEA) in an animal. The method involves priming of the animal
CC   with an inducing agent such as tetanus toxoid (TT) or diphtheria toxoid
CC   (DT), subsequently followed by administration of an inducing agent-
CC   antigen mixture. The method provides the enhancement or augmentation of
CC   the immune response to the antigen and/or improves a vaccination protocol
CC   by allowing use of less antigen. The immunisation of the animal with
CC   tumour-associated antigen is useful for the prophylactic or therapeutic
CC   treatment of cancer. The present DNA sequence encodes modified
CC   carcinoembryonic antigen (CEA) related to the invention
XX
SQ   Sequence 2106 BP; 559 A; 659 C; 442 G; 446 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:                7.06          Length:          2106
Score:                    45.00         Matches:         9
Percent Similarity:       100.00%       Conservative:    0
Best Local Similarity:    100.00%       Mismatches:      0
Query Match:              100.00%       Indels:          0
DB:                       5             Gaps:            0

US-10-725-373-2 (1-9) x AAD07347 (1-2106)

Qy          1 TyrLeuSerGlyAlaAspLeuAsnLeu 9
              |||||||||||||||||||||||||||
Db       1810 TACCTTTCGGGAGCGGACCTCAACCTC 1836

RESULT 5
AAI72489
ID   AAI72489 standard; DNA; 2106 BP.
XX
AC   AAI72489;
XX
DT   16-MAY-2002  (first entry)
XX
DE   CEA agonist coding sequence #1.
XX
KW   CAP-1; epitope; carcinoembryonic antigen; CEA; agonist; immune response;
KW   carcinoma; gastrointestinal; breast; pancreatic; bladder; ovarian; lung;
KW   prostate; cancer; gene; ds.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   CDS             1. .2106
FT                   /*tag=  a
FT                   /product= "CEA agonist polypeptide"
XX
PN   WO200210379-A2.
XX
PD   07-FEB-2002.
XX
PF   27-JUL-2001; 2001WO-CA001092.
XX
PR   31-JUL-2000; 2000US-0222043P.
XX
PA   (AVET ) AVENTIS PASTEUR LTD.
PA   (THER-) THERION BIOLOGICS.
PA   (USSH ) US NAT CANCER INST.
XX
PI   Berinstein N,  Tartaglia J,  Tine JA,  Panicali DL,  Gritz L;
PI   Schlom J;
XX
```

DR WPI; 2002-206189/26.
DR P-PSDB; AAB47918.
XX
PT Carcinoembryonic antigen agonist polypeptide for inducing an immune
PT response in animal against antigen and for inhibiting an epitope antigen
PT expressing carcinoma cell, comprises a modified antigen epitope.
XX
PS Claim 4; Fig 1; 69pp; English.
XX
CC This sequence encodes the carcinoembryonic antigen (CEA) agonist
CC polypeptide of the invention. This sequence represents the sequence given
CC in the figures in the specification, and it directly encodes the CEA
CC agonist polypeptide given in AAB47918. The CEA agonist contains a
CC modified CAP-1 epitope of CEA, in which position 6 is modified from Asp
CC to Asn to increase its immunogenicity. The CEA agonist polypeptide of the
CC invention, or DNA encoding it, are useful for: (i) inducing an immune
CC response in an animal directed against a CEA protein or fragment, CEA
CC agonist, a CEA epitope, a modified CEA epitope, cells expressing or
CC binding a CEA protein or fragment; and (ii) inhibiting a CEA epitope
CC expressing carcinoma cell, which is a gastrointestinal, breast,
CC pancreatic, bladder, ovarian, lung or prostate carcinoma cell in a
CC patient, hence is useful for manufacture of a medicament for the
CC treatment of cancer
XX
SQ Sequence 2106 BP; 559 A; 658 C; 442 G; 447 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	7.06	Length:	2106
Score:	45.00	Matches:	9
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	6	Gaps:	0

US-10-725-373-2 (1-9) x AAI72489 (1-2106)

```
Qy          1 TyrLeuSerGlyAlaAspLeuAsnLeu 9
              |||||||||||||||||||||||||||
Db       1810 TACCTTTCGGGAGCGGACCTCAACCTC 1836
```

RESULT 6
ADE13860
ID ADE13860 standard; DNA; 2106 BP.
XX
AC ADE13860;
XX
DT 29-JAN-2004 (first entry)
XX
DE CEA-CAP6D nucleotide sequence SEQ ID NO:23.
XX
KW carcinoembryonic antigen; CEA; CEA(6D)-1; 2; cytostatic; vaccine; cancer;
KW tumour antigen; immunotherapy; gene; ds.
XX
OS Unidentified.
XX
PN WO2003085087-A2.
XX
PD 16-OCT-2003.
XX
PF 09-APR-2003; 2003WO-US010916.
XX
PR 09-APR-2002; 2002US-0372972P.
XX
PA (AVET) AVENTIS PASTEUR LTD.
PA (THER-) THERION BIOLOGICS INC.
XX
PI Parrington M, Zhang L, Rovinski B, Gritz LR, Greenhalgh T;
XX
DR WPI; 2003-877029/81.
XX
PT New isolated DNA molecule comprising the carcinoembryonic antigen (6D)-
PT 1,2 sequence, useful for diagnosing, preventing and treating cancer, or
PT determining the effectiveness of a chemotherapeutic or other treatment
PT regimen.
XX
PS Example 1; SEQ ID NO 23; 56pp; English.
XX
CC The present invention describes an isolated DNA molecule comprising the
CC carcinoembryonic antigen (CEA) (6D)-1,2 sequence of 2106 bp (see
CC ADE13861), or its fragment. Also described: (1) an expression vector
CC comprising the nucleic acid sequence CEA(6D)-1,2, or its fragment
CC describes above; (2) a composition comprising the expression vector of
CC (1) in a pharmaceutical carrier; and (3) preventing or treating cancer
CC comprising administering to a host the expression vector of (1). CEA(6D)-
CC 1,2 has cytostatic activity, and can be used in vaccines. The CEA(6D)-1,2
CC nucleic acid and target polypeptide are useful for diagnosing, preventing
CC and treating cancer, predicting prognosis, or determining the
CC effectiveness of a chemotherapeutic or other treatment regimen. The
CC expression vector may be used for the insertion and expression of CEA(6D)
CC -1,2 nucleic acid encoding tumour antigens for the immunotherapeutic
CC treatment of cancer. The target polypeptides are useful in generating
CC antibodies used in screening assays or for immunotherapy. The present
CC sequence represents the CEA-CAP6D nucleotide sequence, which is given in
CC comparison with CEA(6D)-1,2 in the exemplification of the present
CC invention.
XX
SQ Sequence 2106 BP; 559 A; 659 C; 442 G; 446 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	7.06	Length:	2106
Score:	45.00	Matches:	9
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	10	Gaps:	0

US-10-725-373-2 (1-9) x ADE13860 (1-2106)

```
Qy          1 TyrLeuSerGlyAlaAspLeuAsnLeu 9
              |||||||||||||||||||||||||||
Db       1810 TACCTTTCGGGAGCGGACCTCAACCTC 1836
```

RESULT 7
ADE13861
ID ADE13861 standard; DNA; 2106 BP.
XX
AC ADE13861;
XX
DT 29-JAN-2004 (first entry)
XX
DE CEA(6D)-1,2 nucleotide sequence SEQ ID NO:24.
XX
KW carcinoembryonic antigen; CEA; CEA(6D)-1; 2; cytostatic; vaccine; cancer;
KW tumour antigen; immunotherapy; gene; ds.
XX
OS Unidentified.
XX
PN WO2003085087-A2.
XX
PD 16-OCT-2003.
XX
PF 09-APR-2003; 2003WO-US010916.
XX
PR 09-APR-2002; 2002US-0372972P.
XX
PA (AVET) AVENTIS PASTEUR LTD.
PA (THER-) THERION BIOLOGICS INC.
XX
PI Parrington M, Zhang L, Rovinski B, Gritz LR, Greenhalgh T;
XX
DR WPI; 2003-877029/81.
XX
PT New isolated DNA molecule comprising the carcinoembryonic antigen (6D)-
PT 1,2 sequence, useful for diagnosing, preventing and treating cancer, or
PT determining the effectiveness of a chemotherapeutic or other treatment
PT regimen.

```
XX
PS   Claim 1; SEQ ID NO 24; 56pp; English.
XX
CC   The present invention describes an isolated DNA molecule comprising the
CC   carcinoembryonic antigen (CEA) (6D)-1,2 sequence of 2106 bp (see
CC   ADE13861), or its fragment. Also described: (1) an expression vector
CC   comprising the nucleic acid sequence CEA(6D)-1,2, or its fragment
CC   describes above; (2) a composition comprising the expression vector of
CC   (1) in a pharmaceutical carrier; and (3) preventing or treating cancer
CC   comprising administering to a host the expression vector of (1). CEA(6D)-
CC   1,2 has cytostatic activity, and can be used in vaccines. The CEA(6D)-1,2
CC   nucleic acid and target polypeptide are useful for diagnosing, preventing
CC   and treating cancer, predicting prognosis, or determining the
CC   effectiveness of a chemotherapeutic or other treatment regimen. The
CC   expression vector may be used for the insertion and expression of CEA(6D)
CC   -1,2 nucleic acid encoding tumour antigens for the immunotherapeutic
CC   treatment of cancer. The target polypeptides are useful in generating
CC   antibodies used in screening assays or for immunotherapy. The present
CC   sequence represents the CEA(6D)-1,2 nucleotide sequence, which is given
CC   in the exemplification of the present invention.
XX
SQ   Sequence 2106 BP; 574 A; 615 C; 435 G; 482 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:                7.06          Length:          2106
Score:                    45.00         Matches:         9
Percent Similarity:       100.00%       Conservative:    0
Best Local Similarity:    100.00%       Mismatches:      0
Query Match:              100.00%       Indels:          0
DB:                       10            Gaps:            0

US-10-725-373-2 (1-9) x ADE13861 (1-2106)

Qy          1 TyrLeuSerGlyAlaAspLeuAsnLeu 9
              |||||||||||||||||||||||||||
Db       1810 TACCTTTCGGGAGCGGACCTCAACCTC 1836

RESULT 8
AAI72490/c
ID   AAI72490 standard; DNA; 7958 BP.
XX
AC   AAI72490;
XX
DT   16-MAY-2002  (first entry)
XX
DE   H6-promoter human CEAmod/42K-promoted B7.1 insertion cassette.
XX
KW   CAP-1; epitope; carcinoembryonic antigen; CEA; agonist; immune response;
KW   carcinoma; gastrointestinal; breast; pancreatic; bladder; ovarian; lung;
KW   prostate; cancer; gene; ds.
XX
OS   Homo sapiens.
OS   Synthetic.
OS   Chimeric.
XX
FH   Key             Location/Qualifiers
FT   misc_feature    423. .827
FT                   /*tag=  a
FT                   /note= "ALVAC's C5 locus left flanking arm"
FT   CDS             complement(903. .3008)
FT                   /*tag=  b
FT                   /product= "CEA agonist peptide"
FT   promoter        complement(3009. .3132)
FT                   /*tag=  c
FT                   /label= Vaccinia_H6_promoter
FT   promoter        3210. .3275
FT                   /*tag=  d
FT                   /label= 42K_promoter
FT   CDS             3276. .4142
FT                   /*tag=  e
FT                   /product= "Human B7.1"
FT   misc_feature    4184. .5722
FT                   /*tag=  f
FT                   /note= "ALVAC's C5 locus right flanking arm#"
XX
PN   WO200210379-A2.
XX
PD   07-FEB-2002.
XX
PF   27-JUL-2001; 2001WO-CA001092.
XX
PR   31-JUL-2000; 2000US-0222043P.
XX
PA   (AVET ) AVENTIS PASTEUR LTD.
PA   (THER-) THERION BIOLOGICS.
PA   (USSH ) US NAT CANCER INST.
XX
PI   Berinstein N,  Tartaglia J,  Tine JA,  Panicali DL,  Gritz L;
PI   Schlom J;
XX
DR   WPI; 2002-206189/26.
XX
PT   Carcinoembryonic antigen agonist polypeptide for inducing an immune
PT   response in animal against antigen and for inhibiting an epitope antigen
PT   expressing carcinoma cell, comprises a modified antigen epitope.
XX
PS   Example 3; Fig 4; 69pp; English.
XX
CC   This sequence represents ALVAC(2)-CEAmod/hB7.1. This is a coding sequence
CC   containing the H6 promoted modified carcinoembryonic antigen (CEA)
CC   agonist polypeptide of the invention. The CEA agonist contains a modified
CC   CAP-1 epitope of CEA, in which position 6 is modified from Asp to Asn to
CC   increase its immunogenicity. The CEA agonist polypeptide of the
CC   invention, or DNA encoding it, are useful for: (i) inducing an immune
CC   response in an animal directed against a CEA protein or fragment, CEA
CC   agonist, a CEA epitope, a modified CEA epitope, cells expressing or
CC   binding a CEA protein or fragment; and (ii) inhibiting a CEA epitope
CC   expressing carcinoma cell, which is a gastrointestinal, breast,
CC   pancreatic, bladder, ovarian, lung or prostate carcinoma cell in a
CC   patient, hence is useful for manufacture of a medicament for the
CC   treatment of cancer
XX
SQ   Sequence 7958 BP; 2096 A; 1720 C; 1858 G; 2284 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:                31.6          Length:          7958
Score:                    45.00         Matches:         9
Percent Similarity:       100.00%       Conservative:    0
Best Local Similarity:    100.00%       Mismatches:      0
Query Match:              100.00%       Indels:          0
DB:                       6             Gaps:            0

US-10-725-373-2 (1-9) x AAI72490 (1-7958)

Qy          1 TyrLeuSerGlyAlaAspLeuAsnLeu 9
              |||||||||||||||||||||||||||
Db       1199 TACCTTTCGGGAGCGGACCTCAACCTC 1173

RESULT 9
AAX56259
ID   AAX56259 standard; DNA; 27 BP.
XX
AC   AAX56259;
XX
DT   20-JUL-1999  (first entry)
XX
DE   Carcinoembryonic antigen peptide agonist encoding DNA SEQ ID NO:8.
XX
KW   Carcinoembryonic antigen; CEA; human; agonist; antagonist;
KW   immune response; carcinoma; gastrointestinal; breast; pancreatic;
KW   bladder; ovarian; lung; prostatic; T cell proliferation; cancer;
KW   adoptive transfer therapy; autoimmune reaction; immunotherapy; ss.
XX
OS   Homo sapiens.
OS   Synthetic.
XX
```

PN WO9919478-A1.
XX
PD 22-APR-1999.
XX
PF 22-SEP-1998; 98WO-US019794.
XX
PR 10-OCT-1997; 97US-0061589P.
XX
PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
XX
PI Schlom J, Barzaga E, Zaremba S;
XX
DR WPI; 1999-326544/27.
XX
PT Peptide agonists and antagonists of carcinoembryonal antigen.
XX
PS Claim 22; Page 20; 72pp; English.
XX
CC The present invention describes peptides (A) that comprise agonists (Ia)
CC or antagonists (Ib) of human carcinoembryonal antigen (CEA). (Ia) are
CC used in vaccines to kill or inhibit carcinoma cells that express CEA or
CC its epitopes, particularly for treating gastrointestinal, breast,
CC pancreatic, bladder, ovarian, lung or prostatic carcinoma. They can also
CC be used to proliferate T cells, e.g. from vaccinated subjects, for use in
CC adoptive transfer therapy. (Ib) are used to inhibit CEA-specific immune
CC responses, e.g. in vaccinated subjects, to prevent an autoimmune reaction
CC to cancer immunotherapy (i.e. to prevent attack on normal but CEA-
CC expressing cells). (Ia) are more active than native sequence (I) and
CC generate a highly specific and systemic anti-CEA response. Cytotoxic T
CC cells generated recognize both (Ia) and native CEA epitopes. The present
CC sequence encodes a specifically claimed example of (Ia)
XX
SQ Sequence 27 BP; 6 A; 9 C; 6 G; 6 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	0.145	Length:	27
Score:	43.00	Matches:	8
Percent Similarity:	100.00%	Conservative:	1
Best Local Similarity:	88.89%	Mismatches:	0
Query Match:	95.56%	Indels:	0
DB:	2	Gaps:	0

US-10-725-373-2 (1-9) x AAX56259 (1-27)

```
Qy          1 TyrLeuSerGlyAlaAspLeuAsnLeu 9
              ||||||||||||||||:::||||||
Db          1 TACCTTTCGGGAGCGGACATCAACCTC 27
```

RESULT 10
AAX56260
ID AAX56260 standard; DNA; 27 BP.
XX
AC AAX56260;
XX
DT 20-JUL-1999 (first entry)
XX
DE Carcinoembryonic antigen peptide agonist CAP-1 encoding DNA SEQ ID NO:6.
XX
KW Carcinoembryonic antigen; CEA; human; agonist; antagonist;
KW immune response; carcinoma; gastrointestinal; breast; pancreatic;
KW bladder; ovarian; lung; prostatic; T cell proliferation; cancer;
KW adoptive transfer therapy; autoimmune reaction; immunotherapy; ss.
XX
OS Homo sapiens.
OS Synthetic.
XX
PN WO9919478-A1.
XX
PD 22-APR-1999.
XX
PF 22-SEP-1998; 98WO-US019794.
XX
PR 10-OCT-1997; 97US-0061589P.
XX
PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
XX
PI Schlom J, Barzaga E, Zaremba S;
XX
DR WPI; 1999-326544/27.
XX
PT Peptide agonists and antagonists of carcinoembryonal antigen.
XX
PS Disclosure; Page 19; 72pp; English.
XX
CC The present invention describes peptides (A) that comprise agonists (Ia)
CC or antagonists (Ib) of human carcinoembryonal antigen (CEA). (Ia) are
CC used in vaccines to kill or inhibit carcinoma cells that express CEA or
CC its epitopes, particularly for treating gastrointestinal, breast,
CC pancreatic, bladder, ovarian, lung or prostatic carcinoma. They can also
CC be used to proliferate T cells, e.g. from vaccinated subjects, for use in
CC adoptive transfer therapy. (Ib) are used to inhibit CEA-specific immune
CC responses, e.g. in vaccinated subjects, to prevent an autoimmune reaction
CC to cancer immunotherapy (i.e. to prevent attack on normal but CEA-
CC expressing cells). (Ia) are more active than native sequence (I) and
CC generate a highly specific and systemic anti-CEA response. Cytotoxic T
CC cells generated recognize both (Ia) and native CEA epitopes. The present
CC sequence encodes a specifically claimed example of (Ia)
XX
SQ Sequence 27 BP; 6 A; 10 C; 5 G; 6 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	0.674	Length:	27
Score:	40.00	Matches:	8
Percent Similarity:	100.00%	Conservative:	1
Best Local Similarity:	88.89%	Mismatches:	0
Query Match:	88.89%	Indels:	0
DB:	2	Gaps:	0

US-10-725-373-2 (1-9) x AAX56260 (1-27)

```
Qy          1 TyrLeuSerGlyAlaAspLeuAsnLeu 9
              ||||||||||||||:::|||||||||
Db          1 TACCTTTCGGGAGCGAACCTCAACCTC 27
```

RESULT 11
ADL46174
ID ADL46174 standard; DNA; 30 BP.
XX
AC ADL46174;
XX
DT 17-JUN-2004 (first entry)
XX
DE Human CAP-1 tumour antigen fragment DNA, SEQ ID NO:8 #1.
XX
KW Tumour antigen; vaccine; immunoglobulin; CH3 fragment; fusion protein;
KW cancer; tumour; dendritic cell; endocytosis; immune response; cytostatic;
KW human; CAP-1; ds.
XX
OS Homo sapiens.
XX
PN WO2004024181-A1.
XX
PD 25-MAR-2004.
XX
PF 15-SEP-2003; 2003WO-CN000776.
XX
PR 13-SEP-2002; 2002CN-00136965.
XX
PA (LIJJ/) LI J.
XX
PI Li J;
XX
DR WPI; 2004-269898/25.
XX
PT Tumor-antigen vaccines with molecular weight far smaller than antigen-
PT antibody compound to enable endocytosis by dendritic cells to promote

```
PT   very high immunoreaction for killing tumor cells.
XX
PS   Example 2; SEQ ID NO 8; 28pp; Chinese.
XX
CC   The invention relates to a tumour antigen vaccine comprising 7 or more
CC   amino acids of a tumour antigen sequence joined to an immunoglobulin CH3
CC   fragment. The invention also relates to DNA sequences encoding the
CC   antigenic fusion polypeptide; expression vectors and host cells
CC   comprising the DNA sequences; a process for recombinantly producing the
CC   fusion polypeptide; and a vaccine composition comprising the fusion
CC   polypeptide and a pharmaceutically acceptable carrier. The tumour antigen
CC   used is preferably selected from 07-AP, AFP, ART-4, BAGE B, beta-
CC   catenin/m, bcr-abl, CAMEL, CAP-1, CASP-8, CDC27m, CDK4/m, CEA, CT, Cyp-B,
CC   DAM, ELF2M, ETV6-AML1, ETS, G250, GAGE, GnT-V, GP100, HAGE, HER- 2/NEU,
CC   HLA-A*0201-R170I, HPV-E6, HPV-E7, EBNA, HSP70-2M, HST-2, hTERT, iCE,
CC   KIAA0205, LAGE, LDLR/FUT, GDP-Lfucose, MAGE, MART-1/Melan-A, MCIR,
CC   Myosin/m, MUC1, MUM-1,-2,-3, NA88-A, NY-ESO-1, P15, pl90, P53, Pml/RAR
CC   alpha, FRAME, PSA, PSM, RAGE, RAS, RU1, RU2, SAGE, SART-1, SART-3,
CC   TEL/AML1, TPI/m, TRP-1, gp75, TRP-2, TRP-2/INT2 and WTI. The tumour
CC   antigen vaccines of the invention are useful in cancer therapy. The
CC   antigenic fusion protein used in the vaccine are much smaller than the
CC   corresponding antibody-antigen complex, permitting them to be endocytosed
CC   by dendritic cells and thereby resulting in a greatly increased anti-
CC   tumour immune response. The present sequence represents DNA encoding a
CC   fragment of the human CAP-1/CEA tumour antigen used in an example of the
CC   invention. Note: The present sequence differs from that also referred to
CC   as SEQ ID NO:8 () which is given on page 10 of the specification.
XX
SQ   Sequence 30 BP; 6 A; 12 C; 5 G; 7 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:                0.759          Length:          30
Score:                    40.00          Matches:         8
Percent Similarity:       100.00%        Conservative:    1
Best Local Similarity:    88.89%         Mismatches:      0
Query Match:              88.89%         Indels:          0
DB:                       12             Gaps:            0

US-10-725-373-2 (1-9) x ADL46174 (1-30)

Qy          1 TyrLeuSerGlyAlaAspLeuAsnLeu 9
              ||||||||||||||:::|||||||||
Db          1 TACCTTTCGGGAGCGAACCTCAACCTC 27

RESULT 12
ADL46175
ID   ADL46175 standard; DNA; 64 BP.
XX
AC   ADL46175;
XX
DT   17-JUN-2004  (first entry)
XX
DE   Human immunoglobulin Fc fragment 5' PCR primer, SEQ ID NO:9 #1.
XX
KW   Tumour antigen; vaccine; immunoglobulin; CH3 fragment; fusion protein;
KW   cancer; tumour; dendritic cell; endocytosis; immune response; cytostatic;
KW   human; Fc fragment; PCR; primer; ss.
XX
OS   Homo sapiens.
XX
PN   WO2004024181-A1.
XX
PD   25-MAR-2004.
XX
PF   15-SEP-2003; 2003WO-CN000776.
XX
PR   13-SEP-2002; 2002CN-00136965.
XX
PA   (LIJJ/) LI J.
XX
PI   Li J;
XX
DR   WPI; 2004-269898/25.
XX
PT   Tumor-antigen vaccines with molecular weight far smaller than antigen-
PT   antibody compound to enable endocytosis by dendritic cells to promote
PT   very high immunoreaction for killing tumor cells.
XX
PS   Example 2; SEQ ID NO 9; 28pp; Chinese.
XX
CC   The invention relates to a tumour antigen vaccine comprising 7 or more
CC   amino acids of a tumour antigen sequence joined to an immunoglobulin CH3
CC   fragment. The invention also relates to DNA sequences encoding the
CC   antigenic fusion polypeptide; expression vectors and host cells
CC   comprising the DNA sequences; a process for recombinantly producing the
CC   fusion polypeptide; and a vaccine composition comprising the fusion
CC   polypeptide and a pharmaceutically acceptable carrier. The tumour antigen
CC   used is preferably selected from 07-AP, AFP, ART-4, BAGE B, beta-
CC   catenin/m, bcr-abl, CAMEL, CAP-1, CASP-8, CDC27m, CDK4/m, CEA, CT, Cyp-B,
CC   DAM, ELF2M, ETV6-AML1, ETS, G250, GAGE, GnT-V, GP100, HAGE, HER- 2/NEU,
CC   HLA-A*0201-R170I, HPV-E6, HPV-E7, EBNA, HSP70-2M, HST-2, hTERT, iCE,
CC   KIAA0205, LAGE, LDLR/FUT, GDP-Lfucose, MAGE, MART-1/Melan-A, MCIR,
CC   Myosin/m, MUC1, MUM-1,-2,-3, NA88-A, NY-ESO-1, P15, pl90, P53, Pml/RAR
CC   alpha, FRAME, PSA, PSM, RAGE, RAS, RU1, RU2, SAGE, SART-1, SART-3,
CC   TEL/AML1, TPI/m, TRP-1, gp75, TRP-2, TRP-2/INT2 and WTI. The tumour
CC   antigen vaccines of the invention are useful in cancer therapy. The
CC   antigenic fusion protein used in the vaccine are much smaller than the
CC   corresponding antibody-antigen complex, permitting them to be endocytosed
CC   by dendritic cells and thereby resulting in a greatly increased anti-
CC   tumour immune response. Sequences ADL46175-ADL46176 represent PCR primers
CC   used to amplify DNA encoding a human immunoglobulin Fc fragment in an
CC   example of the invention. Note: The present sequence differs from that
CC   also referred to as SEQ ID NO:9 () which is given on page 10 of the
CC   specification.
XX
SQ   Sequence 64 BP; 16 A; 22 C; 13 G; 13 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:                1.78           Length:          64
Score:                    40.00          Matches:         8
Percent Similarity:       100.00%        Conservative:    1
Best Local Similarity:    88.89%         Mismatches:      0
Query Match:              88.89%         Indels:          0
DB:                       12             Gaps:            0

US-10-725-373-2 (1-9) x ADL46175 (1-64)

Qy          1 TyrLeuSerGlyAlaAspLeuAsnLeu 9
              ||||||||||||||:::|||||||||
Db         15 TACCTTTCGGGAGCGAACCTCAACCTC 41

RESULT 13
AAV57948/c
ID   AAV57948 standard; DNA; 80 BP.
XX
AC   AAV57948;
XX
DT   24-NOV-1998  (first entry)
XX
DE   708 vkcea primary reaction 1.2 oligonucleotide vkcea592r.
XX
KW   Hepatitis B surface antigen; HBsAg; MHC class II-restricted peptide;
KW   vaccination; vaccine; MHC class I molecule; immune response; cancer;
KW   major histocompatibility complex molecule; pathogenic organism;
KW   viral disease; autoimmune condition; allergy; PCR primer; ss.
XX
OS   Synthetic.
XX
PN   WO9833523-A1.
XX
PD   06-AUG-1998.
XX
PF   02-FEB-1998;   98WO-GB000325.
XX
PR   31-JAN-1997;   97GB-00001999.
PR   05-JUL-1997;   97GB-00014182.
```

PR 07-AUG-1997; 97GB-00016620.
PR 07-AUG-1997; 97GB-00016641.
PR 21-NOV-1997; 97GB-00024584.
XX
PA (BIOV-) BIOVATION LTD.
XX
PI Carr FJ, Carter G;
XX
DR WPI; 1998-437178/37.
XX
PT Immunogenic molecules - comprising nucleic acid and polypeptide portion,
PT from both of which peptide for presentation on major histocompatibility
PT complex molecules can be derived.
XX
PS Example 10; Page 60; 87pp; English.
XX
CC A molecule has been developed which comprises: (a) a nucleic acid portion
CC from which at least one peptide for presentation of MHC class I or class
CC II molecules, or both, may be derived, and (b) a polypeptide portion,
CC from which at least 1 peptide for presentation on MHC class I or class II
CC molecules, or both, may be derived. Also described in the present
CC invention is another molecule comprising: (a) a nucleic acid portion from
CC which at least 1 peptide for presentation on MHC class I or class II
CC molecules, or both, may be derived, and (b) a polypeptide^portion
CC comprising a recognition domain capable of targeting the molecule to an
CC antigen presenting cell (APC), where the polypeptide portion does not
CC comprise a specific antigen binding site. The molecules can be used to
CC induce immune responses to treat or prevent, e.g. diseases caused by
CC pathogenic organisms, cancers, viral disease, e.g. HIV or hepatitis
CC infection, autoimmune conditions, e.g. Grave's disease, multiple
CC sclerosis, systemic lupus erythematosus, diabetes mellitus, Kawasaki's
CC disease, rheumatoid arthritis or allergies, e.g. atopic dermatitis,
CC allergic rhinitis, allergic conjunctivitis, atopic asthma or eczema. The
CC combination of DNA and polypeptide in the same molecule can give rise not
CC only to a combination of MHC class I- and MHC class II-mediated immune
CC responses but also to an enhancement of these responses compared to the
CC responses given by either DNA or polypeptide alone. The present sequence
CC represents an oligonucleotide used in an example from the present
CC invention
XX
SQ Sequence 80 BP; 16 A; 28 C; 19 G; 17 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	2.29	Length:	80
Score:	40.00	Matches:	8
Percent Similarity:	100.00%	Conservative:	1
Best Local Similarity:	88.89%	Mismatches:	0
Query Match:	88.89%	Indels:	0
DB:	2	Gaps:	0

US-10-725-373-2 (1-9) x AAV57948 (1-80)

```
Qy          1 TyrLeuSerGlyAlaAspLeuAsnLeu 9
              ||||||||||||||:::|||||||||
Db         68 TACCTGTCCGGCGCCAACCTGAACCTG 42
```

RESULT 14
AAV81101/c
ID AAV81101 standard; DNA; 80 BP.
XX
AC AAV81101;
XX
DT 03-MAR-1999 (first entry)
XX
DE Vaccine 2 708 Vl constructing long oligo VKCEA592R.
XX
KW Non-immunogenic; epitope; T-cell; immunogenicity; immune system; SK;
KW immunoglobulin; therapeutic; streptokinase; vaccine; 708; ss.
XX
OS Synthetic.
XX
PN WO9852976-A1.
XX
PD 26-NOV-1998.
XX
PF 21-MAY-1998; 98WO-GB001473.
XX
PR 21-MAY-1997; 97GB-00010480.
PR 31-JUL-1997; 97GB-00016197.
PR 28-NOV-1997; 97GB-00025270.
PR 02-DEC-1997; 97US-0067235P.
PR 14-APR-1998; 98GB-00007751.
XX
PA (BIOV-) BIOVATION LTD.
XX
PI Carr FJ;
XX
DR WPI; 1999-045301/04.
XX
PT Reducing immunogenicity of proteins - by modifying the amino acid
PT sequence of the protein to eliminate potential epitopes for T-cells of a
PT given species.
XX
PS Example 4; Fig 20; 77pp; English.
XX
CC The invention relates to a method for the production of non-immunogenic
CC proteins. The method comprises determining at least part of the amino
CC acid sequence of the protein; (b) identifying in the amino acid sequence
CC one or more potential epitopes for T-cells (T-cell epitopes) of the given
CC species; and (c) modifying the amino acid sequence to eliminate at least
CC one of the T-cell epitopes identified in step (b) thereby to eliminate or
CC reduce the immunogenicity of the protein when exposed to the immune
CC system of the given species. A method of analysing a pre-existing protein
CC to predict the basis for immunogenic responses is also provided. The
CC methods can be used particularly for reducing the immunogenicity of
CC immunoglobulins or therapeutic proteins, e.g. Streptokinase (SK). The
CC products can be used for diagnosis and therapy. Sequences AAV81090-110
CC represent oligonucleotides used for the construction of vaccine 2 708 Vh
CC and Vl
XX
SQ Sequence 80 BP; 16 A; 28 C; 19 G; 17 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	2.29	Length:	80
Score:	40.00	Matches:	8
Percent Similarity:	100.00%	Conservative:	1
Best Local Similarity:	88.89%	Mismatches:	0
Query Match:	88.89%	Indels:	0
DB:	2	Gaps:	0

US-10-725-373-2 (1-9) x AAV81101 (1-80)

```
Qy          1 TyrLeuSerGlyAlaAspLeuAsnLeu 9
              ||||||||||||||:::|||||||||
Db         68 TACCTGTCCGGCGCCAACCTGAACCTG 42
```

RESULT 15
AAI29234
ID AAI29234 standard; cDNA; 155 BP.
XX
AC AAI29234;
XX
DT 12-OCT-2001 (first entry)
XX
DE Colon tumour related determined cDNA sequence for clone R0094:D08.
XX
KW Human; immunotherapy; diagnosis; colon cancer; colon tumour; immunogenic;
KW gene therapy; vaccine; colonic cancer; ss.
XX
OS Homo sapiens.
XX
PN WO200149716-A2.
XX
PD 12-JUL-2001.
XX
PF 29-DEC-2000; 2000WO-US035596.

XX
PR 30-DEC-1999; 99US-00476296.
PR 10-JAN-2000; 2000US-00480321.
PR 15-FEB-2000; 2000US-00504629.
PR 06-MAR-2000; 2000US-00519444.
PR 19-MAY-2000; 2000US-00575251.
PR 29-JUN-2000; 2000US-00609448.
PR 28-AUG-2000; 2000US-00649811.
XX
PA (CORI-) CORIXA CORP.
XX
PI Xu J, Lodes MJ, Secrist H, Benson DR, Meagher MJ, Stolk JA;
PI King GE, Wang T, Jiang Y;
XX
DR WPI; 2001-441847/47.
XX
PT Colon tumor associated proteins and nucleic acids useful for the
PT prevention, diagnosis and treatment of colonic cancer.
XX
PS Claim 2; Page 356; 472pp; English.
XX
CC The present invention describes colon tumour associated proteins (I) and
CC the polynucleotides (II) that encode them. (I) have cytostatic activity.
CC (I) and (II) can be used in gene therapy and vaccine production. (I) and
CC (II) may be used in the prevention, diagnosis and treatment of diseases
CC associated with inappropriate colon tumour associated protein (TCAP)
CC expression, such as colonic cancer. For example, (I) and (II) may be used
CC to treat disorders associated with decreased expression by rectifying
CC mutations or deletions in a patient's genome that affect the activity of
CC TCAPs by expressing inactive proteins or to supplement the patients own
CC production of them. Additionally, (II) may be used to produce the TCAP
CC proteins, by inserting the nucleic acids into a host cell culturing the
CC cell to express the protein. (II) and its complementary sequences may
CC also be used as DNA probes in diagnostic polymerase chain reaction (PCR)
CC and hybridisation assays to detect and quantitate the presence of similar
CC nucleic acids in samples, and therefore which patients may be in need of
CC restorative therapy. (I) may also be used as antigens in the production
CC of antibodies against TCAPs and in assays to identify modulators of TCAP
CC expression and activity. Anti-(I) antibodies and antagonists may also be
CC used to down regulate TCAP expression and activity. The anti-(I)
CC antibodies may also be used as diagnostic agents for detecting the
CC presence of TCAPs in samples (e.g. by enzyme linked immunosorbant assay
CC (ELISA)). AAI28460 to AAI29512 and AAM24494 to AAM24523 represent
CC nucleotide and amino acid sequences given in the exemplification of the
CC present invention
XX
SQ Sequence 155 BP; 30 A; 56 C; 31 G; 30 T; 0 U; 8 Other;

Alignment Scores:

Pred. No.:	4.84	Length:	155
Score:	40.00	Matches:	8
Percent Similarity:	100.00%	Conservative:	1
Best Local Similarity:	88.89%	Mismatches:	0
Query Match:	88.89%	Indels:	0
DB:	4	Gaps:	0

US-10-725-373-2 (1-9) x AAI29234 (1-155)

```
Qy          1 TyrLeuSerGlyAlaAspLeuAsnLeu 9
              ||||||||||||||:::|||||||||
Db        119 TACCTTTCNGGAGCGAACCTCAACCTC 145
```

Search completed: May 17, 2005, 17:45:01
Job time : 329.5 secs

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: May 17, 2005, 06:13:14 ; Search time 66 Seconds
(without alignments)
52.740 Million cell updates/sec

Title: US-10-725-373-3
Perfect score: 45
Sequence: 1 YLSGADINL 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_16Dec04:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	45	100.0	9	2	AAY09527	Aay09527 Carcinoem
2	43	95.6	9	2	AAY09526	Aay09526 Carcinoem
3	43	95.6	9	3	AAB13750	Aab13750 Peptide f
4	43	95.6	9	4	AAB97818	Aab97818 Carcinoem
5	43	95.6	9	4	AAE05124	Aae05124 Modified
6	43	95.6	9	5	AAB47917	Aab47917 Modified
7	43	95.6	9	5	AAE19089	Aae19089 HLA-A24 r
8	43	95.6	9	8	ADN63713	Adn63713 HLA bindi
9	43	95.6	701	4	AAB97817	Aab97817 Modified
10	43	95.6	701	4	AAE05117	Aae05117 Modified
11	43	95.6	701	5	AAB47919	Aab47919 CEA agoni
12	43	95.6	701	5	AAB47918	Aab47918 CEA agoni
13	40	88.9	9	2	AAY09528	Aay09528 Carcinoem
14	38	84.4	9	2	AAW39723	Aaw39723 Human car
15	38	84.4	9	2	AAW70045	Aaw70045 CEA deriv
16	38	84.4	9	2	AAW77134	Aaw77134 CEA synth
17	38	84.4	9	2	AAY47655	Aay47655 Immunogen
18	38	84.4	9	2	AAY09525	Aay09525 Carcinoem
19	38	84.4	9	3	AAB13749	Aab13749 Peptide f
20	38	84.4	9	4	AAE02673	Aae02673 Human CEA
21	38	84.4	9	4	AAE00463	Aae00463 Human tum
22	38	84.4	9	4	AAE05123	Aae05123 Carcinoem
23	38	84.4	9	4	AAB82776	Aab82776 Carcinoem
24	38	84.4	9	5	ABG79073	Abg79073 Human CEA
25	38	84.4	9	5	AAE26805	Aae26805 Human HLA
26	38	84.4	9	5	AAU95893	Aau95893 Immunogen
27	38	84.4	9	5	AAE19088	Aae19088 HLA-A24 r
28	38	84.4	9	6	ABR56428	Abr56428 CEA epito
29	38	84.4	9	6	ABP98779	Abp98779 CAE pepti
30	38	84.4	9	6	ABR44529	Abr44529 CEA epito
31	38	84.4	9	7	ADD84715	Add84715 Human car
32	38	84.4	9	7	AAO24210	Aao24210 Human tum
33	38	84.4	9	8	ADG20333	Adg20333 Antigenic
34	38	84.4	9	8	ADJ36382	Adj36382 CEA epito
35	38	84.4	9	8	ADM12344	Adm12344 MHC class
36	38	84.4	9	8	ADM12341	Adm12341 MHC class
37	38	84.4	9	8	ADM72999	Adm72999 Human CEA
38	38	84.4	9	8	ADL46188	Adl46188 Human CAP
39	38	84.4	9	8	ADO38561	Ado38561 Carcinoem
40	38	84.4	9	8	ADO38564	Ado38564 Carcinoem
41	38	84.4	10	2	AAY46555	Aay46555 Immunogen
42	38	84.4	10	5	AAU11587	Aau11587 Human car
43	38	84.4	10	6	ABR83489	Abr83489 Human car
44	38	84.4	10	8	ADM72998	Adm72998 Human CEA
45	38	84.4	10	8	ADP80031	Adp80031 Human HLA

ALIGNMENTS

RESULT 1
AAY09527
ID AAY09527 standard; peptide; 9 AA.
XX
AC AAY09527;
XX
DT 20-JUL-1999 (first entry)
XX
DE Carcinoembryonic antigen peptide agonist SEQ ID NO:3.
XX
KW Carcinoembryonic antigen; CEA; human; agonist; antagonist;
KW immune response; carcinoma; gastrointestinal; breast; pancreatic;
KW bladder; ovarian; lung; prostatic; T cell proliferation; cancer;
KW adoptive transfer therapy; autoimmune reaction; immunotherapy.
XX
OS Homo sapiens.
OS Synthetic.
XX
PN WO9919478-A1.
XX
PD 22-APR-1999.
XX
PF 22-SEP-1998; 98WO-US019794.
XX
PR 10-OCT-1997; 97US-0061589P.
XX
PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
XX
PI Schlom J, Barzaga E, Zaremba S;
XX
DR WPI; 1999-326544/27.
XX
PT Peptide agonists and antagonists of carcinoembryonal antigen.
XX
PS Claim 5; Page 53; 72pp; English.
XX
CC The present invention describes peptides (A) that comprise agonists (Ia)
CC or antagonists (Ib) of human carcinoembryonal antigen (CEA). (Ia) are
CC used in vaccines to kill or inhibit carcinoma cells that express CEA or
CC its epitopes, particularly for treating gastrointestinal, breast,
CC pancreatic, bladder, ovarian, lung or prostatic carcinoma. They can also
CC be used to proliferate T cells, e.g. from vaccinated subjects, for use in
CC adoptive transfer therapy. (Ib) are used to inhibit CEA-specific immune
CC responses, e.g. in vaccinated subjects, to prevent an autoimmune reaction
CC to cancer immunotherapy (i.e. to prevent attack on normal but CEA-
CC expressing cells). (Ia) are more active than native sequence (I) and
CC generate a highly specific and systemic anti-CEA response. Cytotoxic T
CC cells generated recognize both (Ia) and native CEA epitopes. The present

```
CC   sequence represents a specifically claimed example of (Ia)
XX
SQ   Sequence 9 AA;

  Query Match             100.0%;  Score 45;  DB 2;  Length 9;
  Best Local Similarity   100.0%;  Pred. No. 1.8e+06;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 YLSGADINL 9
              |||||||||
Db          1 YLSGADINL 9


RESULT 2
AAY09526
ID   AAY09526 standard; peptide; 9 AA.
XX
AC   AAY09526;
XX
DT   20-JUL-1999  (first entry)
XX
DE   Carcinoembryonic antigen peptide agonist SEQ ID NO:2.
XX
KW   Carcinoembryonic antigen; CEA; human; agonist; antagonist;
KW   immune response; carcinoma; gastrointestinal; breast; pancreatic;
KW   bladder; ovarian; lung; prostatic; T cell proliferation; cancer;
KW   adoptive transfer therapy; autoimmune reaction; immunotherapy.
XX
OS   Homo sapiens.
OS   Synthetic.
XX
PN   WO9919478-A1.
XX
PD   22-APR-1999.
XX
PF   22-SEP-1998;   98WO-US019794.
XX
PR   10-OCT-1997;   97US-0061589P.
XX
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
PI   Schlom J,  Barzaga E,  Zaremba S;
XX
DR   WPI; 1999-326544/27.
XX
PT   Peptide agonists and antagonists of carcinoembryonal antigen.
XX
PS   Claim 5; Page 53; 72pp; English.
XX
CC   The present invention describes peptides (A) that comprise agonists (Ia)
CC   or antagonists (Ib) of human carcinoembryonal antigen (CEA). (Ia) are
CC   used in vaccines to kill or inhibit carcinoma cells that express CEA or
CC   its epitopes, particularly for treating gastrointestinal, breast,
CC   pancreatic, bladder, ovarian, lung or prostatic carcinoma. They can also
CC   be used to proliferate T cells, e.g. from vaccinated subjects, for use in
CC   adoptive transfer therapy. (Ib) are used to inhibit CEA-specific immune
CC   responses, e.g. in vaccinated subjects, to prevent an autoimmune reaction
CC   to cancer immunotherapy (i.e. to prevent attack on normal but CEA-
CC   expressing cells). (Ia) are more active than native sequence (I) and
CC   generate a highly specific and systemic anti-CEA response. Cytotoxic T
CC   cells generated recognize both (Ia) and native CEA epitopes. The present
CC   sequence represents a specifically claimed example of (Ia)
XX
SQ   Sequence 9 AA;

  Query Match             95.6%;  Score 43;  DB 2;  Length 9;
  Best Local Similarity   88.9%;  Pred. No. 1.8e+06;
  Matches    8;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 YLSGADINL 9
              ||||||:||
Db          1 YLSGADLNL 9


RESULT 3
AAB13750
ID   AAB13750 standard; peptide; 9 AA.
XX
AC   AAB13750;
XX
DT   02-FEB-2001  (first entry)
XX
DE   Peptide fragment # 2 from human CEA.
XX
KW   Human; T-cell; immune response; antigen; epitope; B7 family molecule;
KW   Leukocyte function-associated antigen-3; LFA-3;
KW   Intercellular adhesion molecule-1; ICAM-1; vaccine; immunotherapy;
KW   colon polyp; Crohn's disease; ulcerative colitis; breast lesion; tumour;
KW   CEA.
XX
OS   Homo sapiens.
XX
PN   WO200034494-A1.
XX
PD   15-JUN-2000.
XX
PF   12-NOV-1999;   99WO-US026866.
XX
PR   09-DEC-1998;   98US-0111582P.
XX
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICES.
PA   (THER-) THERION BIOLOGICS CORP.
XX
PI   Schlom J,  Hodge J,  Panicali D;
XX
DR   WPI; 2000-431307/37.
XX
PT   Novel recombinant vector useful as immunogens and vaccines for
PT   stimulating and enhancing immunological responses to target cells and
PT   antigens expresses multiple co-stimulatory molecules such as B7-1, LFA-3,
PT   ICAM-1.
XX
PS   Claim 18; Page 35; 188pp; English.
XX
CC   Costimulatory molecules have important roles in T-cell activation and
CC   therefore the immune response. The present invention relates to
CC   recombinant vectors which comprise of foreign nucleic acid sequences
CC   encoding at least three costimulatory molecules: a B7 family molecule,
CC   Leukocyte function-associated antigen-3 (LFA-3, human CD58) and
CC   Intercellular adhesion molecule-1 (ICAM-1, CD54) and optionally a foreign
CC   gene encoding a target antigen or immunological epitope. The present
CC   sequence is one such target antigen used in the present invention. The
CC   present sequence is a tumour-associated antigen. The vector of the
CC   present invention would be useful for providing an enhanced immune
CC   response to the present target antigen. The vector of the present
CC   invention may therefore be useful in immunotherapy for treating or
CC   preventing diseases caused by viruses, bacteria, protozoans, parasites,
CC   premalignant cells and tumour cells. The recombinant vector can be used
CC   to treat or prevent preneoplastic or hyperplastic states such as colon
CC   polyps, Crohn's disease, ulcerative colitis and breast lesions
XX
SQ   Sequence 9 AA;

  Query Match             95.6%;  Score 43;  DB 3;  Length 9;
  Best Local Similarity   88.9%;  Pred. No. 1.8e+06;
  Matches    8;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 YLSGADINL 9
              ||||||:||
Db          1 YLSGADLNL 9


RESULT 4
AAB97818
ID   AAB97818 standard; peptide; 9 AA.
XX
```

AC AAB97818;
XX
DT 08-AUG-2001 (first entry)
XX
DE Carcinoembryonic antigen (CEA) modified antigen SEQ ID NO:113.
XX
KW Virus; adenovirus; poxvirus; alphavirus; immune response; gp100;
KW tumour antigen; CEA; carcinoembryonic antigen; immunostimulant;
KW cytostatic; immunotherapy; interferon-gamma; IFN-gamma; cancer.
XX
OS Unidentified.
XX
PN WO200130382-A1.
XX
PD 03-MAY-2001.
XX
PF 20-OCT-2000; 2000WO-CA001253.
XX
PR 22-OCT-1999; 99US-0160879P.
PR 07-AUG-2000; 2000US-0223325P.
XX
PA (AVET) AVENTIS PASTEUR LTD.
XX
PI Berinstein N, Tartaglia J, Moingeon P, Barber B;
XX
DR WPI; 2001-308587/32.
XX
PT Inducing immune response to tumor antigen, useful in immunotherapy of
PT cancer, by administering the antigen to a lymphatic site.
XX
PS Claim 19; Page 9; 60pp; English.
XX
CC The present invention describes a method for inducing an immune response,
CC in an animal, to a tumour antigen (Ag) comprising administering Ag, or
CC nucleic acid (I) that encodes it, to a lymphatic site. Cynomolgus monkeys
CC (Macaca fascicularis) were injected with a modified form of gp100 antigen
CC (a) into the left inguinal lymph node or (b) subcutaneously. Both animals
CC of (a) developed a cell-mediated response (indicated by production of
CC interferon-gamma from T lymphocytes when exposed to gp100 peptides), but
CC only 2 of 4 animals of (b) did so. Also animals in (a) produced a far
CC greater antibody response to gp100. The method is used in immunotherapy
CC of a wide range of cancers through induction of a specific immune
CC response (humoral and cellular) against the tumour antigens. When
CC administered to a lymphatic site, Ag (or (I)) induces a stronger immune
CC response than administration by other routes and may also break tolerance
CC to Ag. AAB97708 and AAB97709 represent gp100 epitopes; AAB97710 to
CC AAB97815 represent peptides derived from gp100 which stimulate interferon
CC (IFN)-gamma production; AAH20120 encodes the modified gp100 protein given
CC in AAB97816; AAH20121 encodes the modified carcinoembryonic antigen (CEA)
CC protein given in AAB97817; and AAB97818 represents a CEA modified antigen
CC peptide, all of which are used in the exemplification of the present
CC invention
XX
SQ Sequence 9 AA;

Query Match 95.6%; Score 43; DB 4; Length 9;
Best Local Similarity 88.9%; Pred. No. 1.8e+06;
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 YLSGADINL 9
||||||:||
Db 1 YLSGADLNL 9

RESULT 5
AAE05124
ID AAE05124 standard; peptide; 9 AA.
XX
AC AAE05124;
XX
DT 18-SEP-2001 (first entry)
XX
DE Modified carcinoembryonic antigen (CEA) peptide, CAP-6D.
XX
KW Tumour-associated antigen; TAA; cytostatic; vaccine; gene therapy;
KW immune response; tetanus toxoid; TT; diphtheria toxoid; DT; prophylactic;
KW cancer; therapeutic; carcinoembryonic antigen; CEA.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Misc-difference 6
FT /note= "Wild type Asn substituted with Asp"
XX
PN WO200149317-A2.
XX
PD 12-JUL-2001.
XX
PF 05-JAN-2001; 2001WO-CA000005.
XX
PR 05-JAN-2000; 2000US-0174587P.
XX
PA (AVET) AVENTIS PASTEUR LTD.
XX
PI Emtage P, Barber BH, Sambhara S, Sia CDY;
XX
DR WPI; 2001-441790/47.
XX
PT Enhancing immune response to antigen such as tumor antigen for treating
PT cancer in an animal involves administering an inducing agent to the
PT animal followed by administering inducing agent-antigen mixture.
XX
PS Example 2; Page 31; 62pp; English.
XX
CC The invention relates to a method of enhancing an immune response against
CC tumour-associated antigens (TAAs), such as GP100 and carcinoembryonic
CC antigen (CEA) in an animal. The method involves priming of the animal
CC with an inducing agent such as tetanus toxoid (TT) or diphtheria toxoid
CC (DT), subsequently followed by administration of an inducing agent-
CC antigen mixture. The method provides the enhancement or augmentation of
CC the immune response to the antigen and/or improves a vaccination protocol
CC by allowing use of less antigen. The immunisation of the animal with
CC tumour-associated antigen is useful for the prophylactic or therapeutic
CC treatment of cancer. The present sequence is modified carcinoembryonic
CC antigen (CEA) peptide fragment related to the invention
XX
SQ Sequence 9 AA;

Query Match 95.6%; Score 43; DB 4; Length 9;
Best Local Similarity 88.9%; Pred. No. 1.8e+06;
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 YLSGADINL 9
||||||:||
Db 1 YLSGADLNL 9

RESULT 6
AAB47917
ID AAB47917 standard; peptide; 9 AA.
XX
AC AAB47917;
XX
DT 16-MAY-2002 (first entry)
XX
DE Modified CEA epitope, CEA(6D).
XX
KW CAP-1; epitope; carcinoembryonic antigen; CEA; agonist; immune response;
KW carcinoma; gastrointestinal; breast; pancreatic; bladder; ovarian; lung;
KW prostate; cancer.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Misc-difference 6
FT /label= N6D

XX
PN WO200210379-A2.
XX
PD 07-FEB-2002.
XX
PF 27-JUL-2001; 2001WO-CA001092.
XX
PR 31-JUL-2000; 2000US-0222043P.
XX
PA (AVET) AVENTIS PASTEUR LTD.
PA (THER-) THERION BIOLOGICS.
PA (USSH) US NAT CANCER INST.
XX
PI Berinstein N, Tartaglia J, Tine JA, Panicali DL, Gritz L;
PI Schlom J;
XX
DR WPI; 2002-206189/26.
XX
PT Carcinoembryonic antigen agonist polypeptide for inducing an immune
PT response in animal against antigen and for inhibiting an epitope antigen
PT expressing carcinoma cell, comprises a modified antigen epitope.
XX
PS Claim 1; Page 38; 69pp; English.
XX
CC This sequence represents a modified CAP-1 epitope of carcinoembryonic
CC antigen (CEA) which was used as part of the CEA agonist polypeptide of
CC the invention. The modification of position 6 of this peptide from Asp to
CC Asn increases its immunogenicity. The CEA agonist polypeptide of the
CC invention, or DNA encoding it, are useful for: (i) inducing an immune
CC response in an animal directed against a CEA protein or fragment, CEA
CC agonist, a CEA epitope, a modified CEA epitope, cells expressing or
CC binding a CEA protein or fragment; and (ii) inhibiting a CEA epitope
CC expressing carcinoma cell, which is a gastrointestinal, breast,
CC pancreatic, bladder, ovarian, lung or prostate carcinoma cell in a
CC patient, hence is useful for manufacture of a medicament for the
CC treatment of cancer
XX
SQ Sequence 9 AA;

Query Match 95.6%; Score 43; DB 5; Length 9;
Best Local Similarity 88.9%; Pred. No. 1.8e+06;
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 YLSGADINL 9
||||||:||
Db 1 YLSGADLNL 9

RESULT 7
AAE19089
ID AAE19089 standard; peptide; 9 AA.
XX
AC AAE19089;
XX
DT 21-MAY-2002 (first entry)
XX
DE HLA-A24 restricted target antigen CEA immunological epitope #3.
XX
KW Human leukocyte antigen; HLA; pharmaceutical composition; target antigen;
KW immunological epitope; replication-defective virus; RDV; immune response;
KW chemotherapy; granulocyte-monocyte-colony stimulating factor; cytostatic;
KW GM-CSF; MHC; major histocompatibility complex; tumour; head; pancreatic;
KW neck; breast; prostate; colorectal; melanoma; myeloidysplastic syndrome;
KW metastatic breast skin lesion; corticosteroid therapy; erythropoietin;
KW cytopenia; neutropenia; vaccine; immunostimulant.
XX
OS Homo sapiens.
XX
PN WO200195919-A2.
XX
PD 20-DEC-2001.
XX
PF 15-JUN-2001; 2001WO-US019201.
XX
PR 15-JUN-2000; 2000US-0211717P.
XX
PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
PA (THER-) THERION BIOLOGICS CORP.
XX
PI Schlom J, Greiner JW, Kass E, Panicali D;
XX
DR WPI; 2002-205852/26.
XX
PT Composition for enhancing immune responses, particularly anti-tumor
PT responses and treating neutropenia, cytopenia, comprises replication-
PT defective virus encoding granulocyte-monocyte-colony stimulating factor.
XX
PS Claim 9; Page 15; 118pp; English.
XX
CC The present invention relates to a pharmaceutical composition comprising
CC a replication-defective virus (RDV) encoding granulocyte-monocyte-colony
CC stimulating factor (GM-CSF). The invention is useful for enhancing cell-
CC mediated or humoral immune response in an individual, by enhancing
CC migration of APC expressing CD11c^+/I-Ab^+, major histocompatibility
CC complex (MHC) class II, at an injection site, regional lymph node at a
CC tumour site, APC proliferation or function, CD4^+T or CD8^+T cell
CC activation, interleukin (IL)-2, interferon (IFN)-gamma or tumour necrosis
CC factor (TNF)-alpha production or their combinations. The composition
CC enhances an antigen-specific T-cell response in an individual to a target
CC antigen or its immunological epitope and an anti-tumour response in an
CC individual with a head tumour, neck, breast, pancreatic, prostate,
CC colorectal or metastatic tumour or melanoma, or metastatic breast skin
CC lesion. The invention is further useful for treating neutropenia
CC resulting from chemotherapy, corticosteroid therapy, irradiation or an
CC infection, by raising the neutrophil count to normal levels and for
CC treating cytopenias in patients with myeloidysplastic syndrome in
CC combination with erythropoietin, by increasing neutrophil count and
CC erythroid precursors. The composition enhances immune response to
CC vaccines such as DPT, Td, DtaP, Hib, DtaP-Hib, MMR, Hepatitis A,
CC hepatitis B, lyme's disease, influenza, tetravalent meningococcal
CC polysaccharide, pneumococcal polysaccharide, anthrax, cholera, plague,
CC yellow fever and Bacillus Calmette-Guerin vaccine. The present sequence
CC is human leukocyte antigen (HLA)-restricted target tumour antigen
CC immunological epitope
XX
SQ Sequence 9 AA;

Query Match 95.6%; Score 43; DB 5; Length 9;
Best Local Similarity 88.9%; Pred. No. 1.8e+06;
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 YLSGADINL 9
||||||:||
Db 1 YLSGADLNL 9

RESULT 8
ADN63713
ID ADN63713 standard; peptide; 9 AA.
XX
AC ADN63713;
XX
DT 01-JUL-2004 (first entry)
XX
DE HLA binding peptide #313.
XX
KW cytostatic; hepatotropic; virucide; antiinflammatory; anti-HIV;
KW gene therapy; vaccine; HLA binding peptide; HTL epitope; liposome;
KW prostate specific antigen; prostate specific membrane antigen;
KW hepatitis B virus antigen; hepatitis C virus antigen;
KW malignant melanoma antigen; MAGE; Epstein Barr virus; cancer;
KW prostate cancer; AIDS; renal carcinoma; cervical carcinoma; lymphoma;
KW chondyloma acuminatum.
XX
OS Unidentified.
XX

PN WO2004031211-A2.
XX
PD 15-APR-2004.
XX
PF 03-OCT-2003; 2003WO-US031308.
XX
PR 03-OCT-2002; 2002US-0416207P.
PR 08-OCT-2002; 2002US-0417269P.
XX
PA (EPIM-) EPIMMUNE INC.
XX
PI Sidney J, Southwood S, Sette A;
XX
DR WPI; 2004-347953/32.
XX
PT New composition of peptides and nucleic acids capable of binding Major
PT Histocompatibility Complex molecules, useful for diagnosing, preventing
PT or treating viral infections or cancer, such as prostate cancer,
PT hepatitis B or AIDS.
XX
PS Claim 1; SEQ ID NO 313; 186pp; English.
XX
CC The invention relates to a novel composition comprising one or more
CC peptides or nucleic acids encoding an HLA binding peptide. The
CC composition further comprises an HTL epitope. It also comprises a spacer
CC molecule, a carrier, an MHC targeting sequence or a lipid. The peptides
CC are incorporated as part of a liposome. The peptide is from an antigen
CC selected from prostate specific antigen (PSA), prostate specific membrane
CC antigen (PSM), hepatitis B virus (HBV) antigen, hepatitis C virus (HCV)
CC antigen, malignant melanoma antigen (MAGE), Epstein Barr virus, human
CC immunodeficiency type-1 (HIV-I), human immunodeficiency type-2 (HIV-2),
CC Papilloma virus, Lassa virus, Mycobacterium tuberculosis (MT), p53,
CC murine p53 (mp53), CEA, HER2/neu, and tyrosine kinase related protein
CC (TKP). The composition is useful for preventing or treating viral
CC infections or cancer, such as prostate cancer, hepatitis B, hepatitis C,
CC AIDS, renal carcinoma, cervical carcinoma, lymphoma, CMV or chondyloma
CC acuminatum. The composition is also be used for diagnosing such diseases.
CC This sequence represents a peptide of the invention.
XX
SQ Sequence 9 AA;

Query Match 95.6%; Score 43; DB 8; Length 9;
Best Local Similarity 88.9%; Pred. No. 1.8e+06;
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 YLSGADINL 9
||||||:||
Db 1 YLSGADLNL 9

RESULT 9
AAB97817
ID AAB97817 standard; protein; 701 AA.
XX
AC AAB97817;
XX
DT 08-AUG-2001 (first entry)
XX
DE Modified Carcinoembryonic antigen (CEA) protein SEQ ID NO:112.
XX
KW Virus; adenovirus; poxvirus; alphavirus; immune response; gp100;
KW tumour antigen; CEA; carcinoembryonic antigen; immunostimulant;
KW cytostatic; immunotherapy; interferon-gamma; IFN-gamma; cancer.
XX
OS Unidentified.
XX
PN WO200130382-A1.
XX
PD 03-MAY-2001.
XX
PF 20-OCT-2000; 2000WO-CA001253.
XX
PR 22-OCT-1999; 99US-0160879P.
PR 07-AUG-2000; 2000US-0223325P.
XX
PA (AVET) AVENTIS PASTEUR LTD.
XX
PI Berinstein N, Tartaglia J, Moingeon P, Barber B;
XX
DR WPI; 2001-308587/32.
DR N-PSDB; AAH20121.
XX
PT Inducing immune response to tumor antigen, useful in immunotherapy of
PT cancer, by administering the antigen to a lymphatic site.
XX
PS Claim 19; Fig 8; 60pp; English.
XX
CC The present invention describes a method for inducing an immune response,
CC in an animal, to a tumour antigen (Ag) comprising administering Ag, or
CC nucleic acid (I) that encodes it, to a lymphatic site. Cynomolgus monkeys
CC (Macaca fascicularis) were injected with a modified form of gp100 antigen
CC (a) into the left inguinal lymph node or (b) subcutaneously. Both animals
CC of (a) developed a cell-mediated response (indicated by production of
CC interferon-gamma from T lymphocytes when exposed to gp100 peptides), but
CC only 2 of 4 animals of (b) did so. Also animals in (a) produced a far
CC greater antibody response to gp100. The method is used in immunotherapy
CC of a wide range of cancers through induction of a specific immune
CC response (humoral and cellular) against the tumour antigens. When
CC administered to a lymphatic site, Ag (or (I)) induces a stronger immune
CC responsé than administration by other routes and may also break tolerance
CC to Ag. AAB97708 and AAB97709 represent gp100 epitopes; AAB97710 to
CC AAB97815 represent peptides derived from gp100 which stimulate interferon
CC (IFN)-gamma production; AAH20120 encodes the modified gp100 protein given
CC in AAB97816; AAH20121 encodes the modified carcinoembryonic antigen (CEA)
CC protein given in AAB97817; and AAB97818 represents a CEA modified antigen
CC peptide, all of which are used in the exemplification of the present
CC invention
XX
SQ Sequence 701 AA;

Query Match 95.6%; Score 43; DB 4; Length 701;
Best Local Similarity 88.9%; Pred. No. 4.5;
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 YLSGADINL 9
||||||:||
Db 604 YLSGADLNL 612

RESULT 10
AAE05117
ID AAE05117 standard; protein; 701 AA.
XX
AC AAE05117;
XX
DT 18-SEP-2001 (first entry)
XX
DE Modified carcinoembryonic antigen (CEA).
XX
KW Tumour-associated antigen; TAA; cytostatic; vaccine; gene therapy;
KW immune response; tetanus toxoid; TT; diphtheria toxoid; DT; prophylactic;
KW cancer; therapeutic; carcinoembryonic antigen; CEA.
XX
OS Synthetic.
XX
PN WO200149317-A2.
XX
PD 12-JUL-2001.
XX
PF 05-JAN-2001; 2001WO-CA000005.
XX
PR 05-JAN-2000; 2000US-0174587P.
XX
PA (AVET) AVENTIS PASTEUR LTD.
XX
PI Emtage P, Barber BH, Sambhara S, Sia CDY;

XX
DR WPI; 2001-441790/47.
DR N-PSDB; AAD07347.
XX
PT Enhancing immune response to antigen such as tumor antigen for treating
PT cancer in an animal involves administering an inducing agent to the
PT animal followed by administering inducing agent-antigen mixture.
XX
PS Claim 9; Fig 3; 62pp; English.
XX
CC The invention relates to a method of enhancing an immune response against
CC tumour-associated antigens (TAAs), such as GP100 and carcinoembryonic
CC antigen (CEA) in an animal. The method involves priming of the animal
CC with an inducing agent such as tetanus toxoid (TT) or diphtheria toxoid
CC (DT), subsequently followed by administration of an inducing agent-
CC antigen mixture. The method provides the enhancement or augmentation of
CC the immune response to the antigen and/or improves a vaccination protocol
CC by allowing use of less antigen. The immunisation of the animal with
CC tumour-associated antigen is useful for the prophylactic or therapeutic
CC treatment of cancer. The present sequence is modified carcinoembryonic
CC antigen (CEA) related to the invention
XX
SQ Sequence 701 AA;

Query Match 95.6%; Score 43; DB 4; Length 701;
Best Local Similarity 88.9%; Pred. No. 4.5;
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 YLSGADINL 9
||||||:||
Db 604 YLSGADLNL 612

RESULT 11
AAB47919
ID AAB47919 standard; protein; 701 AA.
XX
AC AAB47919;
XX
DT 16-MAY-2002 (first entry)
XX
DE CEA agonist #2.
XX
KW CAP-1; epitope; carcinoembryonic antigen; CEA; agonist; immune response;
KW carcinoma; gastrointestinal; breast; pancreatic; bladder; ovarian; lung;
KW prostate; cancer.
XX
OS Synthetic.
XX
PN WO200210379-A2.
XX
PD 07-FEB-2002.
XX
PF 27-JUL-2001; 2001WO-CA001092.
XX
PR 31-JUL-2000; 2000US-0222043P.
XX
PA (AVET) AVENTIS PASTEUR LTD.
PA (THER-) THERION BIOLOGICS.
PA (USSH) US NAT CANCER INST.
XX
PI Berinstein N, Tartaglia J, Tine JA, Panicali DL, Gritz L;
PI Schlom J;
XX
DR WPI; 2002-206189/26.
XX
PT Carcinoembryonic antigen agonist polypeptide for inducing an immune
PT response in animal against antigen and for inhibiting an epitope antigen
PT expressing carcinoma cell, comprises a modified antigen epitope.
XX
PS Claim 2; Page 64-66; 69pp; English.
XX
CC This sequence shows the carcinoembryonic antigen (CEA) agonist
CC polypeptide of the invention. This sequence represents the sequence given
CC in the Seq ID listing in the specification, and is not directly encoded
CC by the coding sequence given in AAI72497 . The CEA agonist contains a
CC modified CAP-1 epitope of CEA, in which position 6 is modified from Asp
CC to Asn to increase its immunogenicity. The CEA agonist polypeptide of the
CC invention, or DNA encoding it, are useful for: (i) inducing an immune
CC response in an animal directed against a CEA protein or fragment, CEA
CC agonist, a CEA epitope, a modified CEA epitope, cells expressing or
CC binding a CEA protein or fragment; and (ii) inhibiting a CEA epitope
CC expressing carcinoma cell, which is a gastrointestinal, breast,
CC pancreatic, bladder, ovarian, lung or prostate carcinoma cell in a
CC patient, hence is useful for manufacture of a medicament for the
CC treatment of cancer
XX
SQ Sequence 701 AA;

Query Match 95.6%; Score 43; DB 5; Length 701;
Best Local Similarity 88.9%; Pred. No. 4.5;
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 YLSGADINL 9
||||||:||
Db 604 YLSGADLNL 612

RESULT 12
AAB47918
ID AAB47918 standard; protein; 701 AA.
XX
AC AAB47918;
XX
DT 16-MAY-2002 (first entry)
XX
DE CEA agonist #1.
XX
KW CAP-1; epitope; carcinoembryonic antigen; CEA; agonist; immune response;
KW carcinoma; gastrointestinal; breast; pancreatic; bladder; ovarian; lung;
KW prostate; cancer.
XX
OS Synthetic.
XX
PN WO200210379-A2.
XX
PD 07-FEB-2002.
XX
PF 27-JUL-2001; 2001WO-CA001092.
XX
PR 31-JUL-2000; 2000US-0222043P.
XX
PA (AVET) AVENTIS PASTEUR LTD.
PA (THER-) THERION BIOLOGICS.
PA (USSH) US NAT CANCER INST.
XX
PI Berinstein N, Tartaglia J, Tine JA, Panicali DL, Gritz L;
PI Schlom J;
XX
DR WPI; 2002-206189/26.
DR N-PSDB; AAI72489.
XX
PT Carcinoembryonic antigen agonist polypeptide for inducing an immune
PT response in animal against antigen and for inhibiting an epitope antigen
PT expressing carcinoma cell, comprises a modified antigen epitope.
XX
PS Claim 2; Fig 1; 69pp; English.
XX
CC This sequence shows the carcinoembryonic antigen (CEA) agonist
CC polypeptide of the invention. This sequence represents the sequence given
CC in the figures in the specification, and is directly encoded by the
CC coding sequence given in AAI72489. The CEA agonist contains a modified
CC CAP-1 epitope of CEA, in which position 6 is modified from Asp to Asn to
CC increase its immunogenicity. The CEA agonist polypeptide of the
CC invention, or DNA encoding it, are useful for: (i) inducing an immune
CC response in an animal directed against a CEA protein or fragment, CEA

```
CC   agonist, a CEA epitope, a modified CEA epitope, cells expressing or
CC   binding a CEA protein or fragment; and (ii) inhibiting a CEA epitope
CC   expressing carcinoma cell, which is a gastrointestinal, breast,
CC   pancreatic, bladder, ovarian, lung or prostate carcinoma cell in a
CC   patient, hence is useful for manufacture of a medicament for the
CC   treatment of cancer
XX
SQ   Sequence 701 AA;

  Query Match             95.6%;  Score 43;  DB 5;  Length 701;
  Best Local Similarity   88.9%;  Pred. No. 4.5;
  Matches    8;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 YLSGADINL 9
              ||||||:||
Db        604 YLSGADLNL 612


RESULT 13
AAY09528
ID   AAY09528 standard; peptide; 9 AA.
XX
AC   AAY09528;
XX
DT   20-JUL-1999  (first entry)
XX
DE   Carcinoembryonic antigen peptide agonist SEQ ID NO:4.
XX
KW   Carcinoembryonic antigen; CEA; human; agonist; antagonist;
KW   immune response; carcinoma; gastrointestinal; breast; pancreatic;
KW   bladder; ovarian; lung; prostatic; T cell proliferation; cancer;
KW   adoptive transfer therapy; autoimmune reaction; immunotherapy.
XX
OS   Homo sapiens.
OS   Synthetic.
XX
PN   WO9919478-A1.
XX
PD   22-APR-1999.
XX
PF   22-SEP-1998;   98WO-US019794.
XX
PR   10-OCT-1997;   97US-0061589P.
XX
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
PI   Schlom J,  Barzaga E,  Zaremba S;
XX
DR   WPI; 1999-326544/27.
XX
PT   Peptide agonists and antagonists of carcinoembryonal antigen.
XX
PS   Claim 5; Page 53; 72pp; English.
XX
CC   The present invention describes peptides (A) that comprise agonists (Ia)
CC   or antagonists (Ib) of human carcinoembryonal antigen (CEA). (Ia) are
CC   used in vaccines to kill or inhibit carcinoma cells that express CEA or
CC   its epitopes, particularly for treating gastrointestinal, breast,
CC   pancreatic, bladder, ovarian, lung or prostatic carcinoma. They can also
CC   be used to proliferate T cells, e.g. from vaccinated subjects, for use in
CC   adoptive transfer therapy. (Ib) are used to inhibit CEA-specific immune
CC   responses, e.g. in vaccinated subjects, to prevent an autoimmune reaction
CC   to cancer immunotherapy (i.e. to prevent attack on normal but CEA-
CC   expressing cells). (Ia) are more active than native sequence (I) and
CC   generate a highly specific and systemic anti-CEA response. Cytotoxic T
CC   cells generated recognize both (Ia) and native CEA epitopes. The present
CC   sequence represents a specifically claimed example of (Ia)
XX
SQ   Sequence 9 AA;

  Query Match             88.9%;  Score 40;  DB 2;  Length 9;
  Best Local Similarity   88.9%;  Pred. No. 1.8e+06;
  Matches    8;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 YLSGADINL 9
              |||||:|||
Db          1 YLSGANINL 9


RESULT 14
AAW39723
ID   AAW39723 standard; peptide; 9 AA.
XX
AC   AAW39723;
XX
DT   11-JUN-1998  (first entry)
XX
DE   Human carcina-embryonic antigen (CEA) peptide (pos. 571-579).
XX
KW   T cell epitope; immune response; human leukocyte antigen; HLA Class I;
KW   vaccine; immunogenic; major histocompatablilty complex; MHC; B cell;
KW   disease; anti-tumour; anti-viral.
XX
OS   Homo sapiens.
XX
PN   WO9741440-A1.
XX
PD   06-NOV-1997.
XX
PF   28-APR-1997;   97WO-NL000229.
XX
PR   26-APR-1996;   96EP-00201145.
PR   23-DEC-1996;   96EP-00203670.
XX
PA   (UYLE-) RIJKSUNIV LEIDEN.
PA   (SCIS-) SCI SEED CAPITAL INVESTMENTS BV.
XX
PI   Van Der Burg SH,  Kast WM,  Toes REM,  Offringa R,  Melief CJM;
XX
DR   WPI; 1997-549891/50.
XX
PT   Method of selecting T cell peptide epitope(s) - by measuring the
PT   stability of HLA class I-peptide complexes on intact B cells.
XX
PS   Example 3; Page 85; 109pp; English.
XX
CC   Peptides AAW39430-W39734 are used in a novel method for the selection of
CC   immunogenic T-cell peptide epitopes present in polypeptide antigens. The
CC   method involves the identification of peptide sequences capable of
CC   binding to an HLA (human leukocyte antigen) class I molecule and
CC   measuring the binding of this epitope peptide to the HLA class I peptide.
CC   The stability of binding of the peptide and MHC (major histocompatablilty
CC   complex) class I molecule is measured on intact human B cells carrying
CC   the MHC molecule at their cell surfaces. The method can be used to select
CC   peptide epitopes for generating vaccines against a disease associated
CC   with the polypeptide, e.g. cancers or AIDS. The peptide epitopes are
CC   especially T-cell peptide epitopes with strong anti-tumour and anti-viral
CC   immune responses. Peptide AAW39723 is derived from the human carcino-
CC   embryonic antigen (CEA) and has the ability to bind to the human MHC
CC   Class I allele HLA-A2.1
XX
SQ   Sequence 9 AA;

  Query Match             84.4%;  Score 38;  DB 2;  Length 9;
  Best Local Similarity   77.8%;  Pred. No. 1.8e+06;
  Matches    7;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 YLSGADINL 9
              |||||::||
Db          1 YLSGANLNL 9


RESULT 15
AAW70045
ID   AAW70045 standard; peptide; 9 AA.
XX
```

AC AAW70045;
XX
DT 22-OCT-1998 (first entry)
XX
DE CEA derived HLA-A2.1 binding peptide 2 (residues 605-613).
XX
KW Cytotoxic T lymphocyte; CTL; major histocompatibility complex; MHC;
KW human leukocyte antigen; HLA; tumour associated antigen; cancer;
KW antigen presenting cell; APC; immunogenic peptide; immune disorder;
KW viral infection; AIDS; hepatitis; bacterial infection; malaria; CEA;
KW fungal infection; tuberculosis; melanoma; carcinoembryonic antigen.
XX
OS Synthetic.
OS Homo sapiens.
XX
PN WO9833888-A1.
XX
PD 06-AUG-1998.
XX
PF 30-JAN-1998; 98WO-US001959.
XX
PR 31-JAN-1997; 97US-0036696P.
XX
PA (EPIM-) EPIMMUNE INC.
XX
PI Tsai V, Southwood S, Sidney J, Sette A, Celis E;
XX
DR WPI; 1998-437445/37.
XX
PT Production of antigen-specific cytotoxic T cells - by incubating
PT immunogenic peptide(s) from antigen that binds class I major
PT histocompatibility complex molecules with pre-treated antigen presenting
PT cells.
XX
PS Example 6; Page 75; 104pp; English.
XX
CC Sequences shown in AAW70044 to AAW70052 represent peptides derived from
CC carcinoembryonic antigen (CEA). The peptides can bind to a human
CC leukocyte antigen (HLA), HLA-A2.1 and are used to exemplify the method of
CC invention of producing antigen-specific cytotoxic T cells (CTLs) in
CC vitro. The method comprises contacting immunogenic peptides from an
CC antigen that binds class I major histocompatibility complex (MHC)
CC molecules with antigen presenting cells (APCs) pretreated with
CC pretreatment growth factors, and incubating the APCs with purified CD8
CC cells in the presence of at least 2 incubation growth factors, thereby
CC producing antigen-specific CTLs. A method for specifically killing target
CC cells in a human patient is also provided which comprises obtaining a
CC fluid sample containing CTLs from a patient, contacting the cytotoxic T
CC cells with APCs pretreated with pre-treatment growth factors, where the
CC APCs comprise class I MHC molecules. The pretreated APCs are incubated
CC with the cytotoxic growth factors, thereby producing activated CTLs which
CC are contacted with a carrier to form a composition. The composition can
CC then be administered to the patient. The activated CTLs can be used for
CC treating cancers, immune disorders, viral infections, AIDS, hepatitis,
CC bacterial infection, fungal infection, malaria or tuberculosis
XX
SQ Sequence 9 AA;

Query Match 84.4%; Score 38; DB 2; Length 9;
Best Local Similarity 77.8%; Pred. No. 1.8e+06;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 YLSGADINL 9
              |||||::||
Db          1 YLSGANLNL 9
```

Search completed: May 17, 2005, 06:17:50
Job time : 67 secs

GenCore version 5.1.6

OM protein - nucleic search, using frame_plus_p2n model

Run on: May 17, 2005, 16:29:39 ; Search time 326.5 Seconds
(without alignments)
163.178 Million cell updates/sec

Title: US-10-725-373-3
Perfect score: 45
Sequence: 1 YLSGADINL 9

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/US10725373/runat_17052005_071020_16177/app_query.fasta_1.796
-DB=N_Geneseq_16Dec04 -QFMT=fastap -SUFFIX=rng -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10725373_@CGN_1_1_1241_@runat_17052005_071020_16177 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N_Geneseq_16Dec04:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002as:*
7: geneseqn2002bs:*
8: geneseqn2003as:*
9: geneseqn2003bs:*
10: geneseqn2003cs:*
11: geneseqn2003ds:*
12: geneseqn2004as:*
13: geneseqn2004bs:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	45	100.0	27	2	AAX56259	Aax56259 Carcinoem
2	43	95.6	27	2	AAX56258	Aax56258 Carcinoem
3	43	95.6	2105	6	AAI72497	Aai72497 CEA agoni
4	43	95.6	2106	4	AAH20121	Aah20121 Modified
5	43	95.6	2106	5	AAD07347	Aad07347 Modified
6	43	95.6	2106	6	AAI72489	Aai72489 CEA agoni
7	43	95.6	2106	10	ADE13860	Ade13860 CEA-CAP6D
8	43	95.6	2106	10	ADE13861	Ade13861 CEA(6D)-1
c 9	43	95.6	7958	6	AAI72490	Aai72490 H6-promot
c 10	40	88.9	108845	13	ABD32542	Abd32542 Mouse can
c 11	39	86.7	2000	6	ABZ16916	Abz16916 Arabidops
12	38	84.4	27	2	AAX56260	Aax56260 Carcinoem
13	38	84.4	30	12	ADL46174	Adl46174 Human CAP
14	38	84.4	64	12	ADL46175	Adl46175 Human imm
c 15	38	84.4	80	2	AAV57948	Aav57948 708 vkcea
c 16	38	84.4	80	2	AAV81101	Aav81101 Vaccine 2
17	38	84.4	155	4	AAI29234	Aai29234 Colon tum
18	38	84.4	155	8	ABZ33420	Abz33420 Human col
19	38	84.4	256	4	AAS57750	Aas57750 cDNA #426
c 20	38	84.4	340	6	ABV88334	Abv88334 Human col
c 21	38	84.4	407	4	AAS57366	Aas57366 cDNA #42
c 22	38	84.4	409	4	AAS57425	Aas57425 cDNA #101
23	38	84.4	409	6	ABV86774	Abv86774 Human col
c 24	38	84.4	409	6	ABV87551	Abv87551 Human col
25	38	84.4	409	6	ABV89100	Abv89100 Human col
26	38	84.4	409	6	ABV87855	Abv87855 Human col
27	38	84.4	409	6	ABK39290	Abk39290 DNA encod
28	38	84.4	409	6	ABK39002	Abk39002 cDNA enco
c 29	38	84.4	409	6	ABK39424	Abk39424 DNA encod
30	38	84.4	409	6	ABK45946	Abk45946 cDNA enco
31	38	84.4	409	6	ABK45289	Abk45289 cDNA enco
c 32	38	84.4	409	6	ABK27782	Abk27782 Human col
33	38	84.4	409	8	ACA11619	Aca11619 Human lun
34	38	84.4	409	8	ACA11331	Aca11331 Human lun
c 35	38	84.4	409	8	ACA11753	Aca11753 Human lun
c 36	38	84.4	409	8	ACA02939	Aca02939 Lung canc
37	38	84.4	409	8	ACA02805	Aca02805 Lung canc
38	38	84.4	409	8	ACA02517	Aca02517 Lung canc
39	38	84.4	409	10	ADH46559	Adh46559 Human lun
40	38	84.4	409	10	ADH46847	Adh46847 Human lun
c 41	38	84.4	409	10	ADH46981	Adh46981 Human lun
42	38	84.4	409	13	ADJ20766	Adj20766 Human lun
43	38	84.4	409	13	ADJ20478	Adj20478 Human lun
c 44	38	84.4	409	13	ADJ20900	Adj20900 Human lun
c 45	38	84.4	410	4	AAS57430	Aas57430 cDNA #106

ALIGNMENTS

RESULT 1
AAX56259
ID AAX56259 standard; DNA; 27 BP.
XX
AC AAX56259;
XX
DT 20-JUL-1999 (first entry)
XX
DE Carcinoembryonic antigen peptide agonist encoding DNA SEQ ID NO:8.
XX
KW Carcinoembryonic antigen; CEA; human; agonist; antagonist;
KW immune response; carcinoma; gastrointestinal; breast; pancreatic;
KW bladder; ovarian; lung; prostatic; T cell proliferation; cancer;
KW adoptive transfer therapy; autoimmune reaction; immunotherapy; ss.
XX
OS Homo sapiens.
OS Synthetic.
XX
PN WO9919478-A1.
XX
PD 22-APR-1999.
XX
PF 22-SEP-1998; 98WO-US019794.
XX
PR 10-OCT-1997; 97US-0061589P.
XX
PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
XX
PI Schlom J, Barzaga E, Zaremba S;

XX
DR WPI; 1999-326544/27.
XX
PT Peptide agonists and antagonists of carcinoembryonal antigen.
XX
PS Claim 22; Page 20; 72pp; English.
XX
CC The present invention describes peptides (A) that comprise agonists (Ia)
CC or antagonists (Ib) of human carcinoembryonal antigen (CEA). (Ia) are
CC used in vaccines to kill or inhibit carcinoma cells that express CEA or
CC its epitopes, particularly for treating gastrointestinal, breast,
CC pancreatic, bladder, ovarian, lung or prostatic carcinoma. They can also
CC be used to proliferate T cells, e.g. from vaccinated subjects, for use in
CC adoptive transfer therapy. (Ib) are used to inhibit CEA-specific immune
CC responses, e.g. in vaccinated subjects, to prevent an autoimmune reaction
CC to cancer immunotherapy (i.e. to prevent attack on normal but CEA-
CC expressing cells). (Ia) are more active than native sequence (I) and
CC generate a highly specific and systemic anti-CEA response. Cytotoxic T
CC cells generated recognize both (Ia) and native CEA epitopes. The present
CC sequence encodes a specifically claimed example of (Ia)
XX
SQ Sequence 27 BP; 6 A; 9 C; 6 G; 6 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	0.0331	Length:	27
Score:	45.00	Matches:	9
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	2	Gaps:	0

US-10-725-373-3 (1-9) x AAX56259 (1-27)

```
Qy          1 TyrLeuSerGlyAlaAspIleAsnLeu 9
              |||||||||||||||||||||||||||
Db          1 TACCTTTCGGGAGCGGACATCAACCTC 27
```

RESULT 2
AAX56258
ID AAX56258 standard; DNA; 27 BP.
XX
AC AAX56258;
XX
DT 20-JUL-1999 (first entry)
XX
DE Carcinoembryonic antigen peptide agonist encoding DNA SEQ ID NO:7.
XX
KW Carcinoembryonic antigen; CEA; human; agonist; antagonist;
KW immune response; carcinoma; gastrointestinal; breast; pancreatic;
KW bladder; ovarian; lung; prostatic; T cell proliferation; cancer;
KW adoptive transfer therapy; autoimmune reaction; immunotherapy; ss.
XX
OS Homo sapiens.
OS Synthetic.
XX
PN WO9919478-A1.
XX
PD 22-APR-1999.
XX
PF 22-SEP-1998; 98WO-US019794.
XX
PR 10-OCT-1997; 97US-0061589P.
XX
PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
XX
PI Schlom J, Barzaga E, Zaremba S;
XX
DR WPI; 1999-326544/27.
XX
PT Peptide agonists and antagonists of carcinoembryonal antigen.
XX
PS Claim 22; Page 20; 72pp; English.
XX
CC The present invention describes peptides (A) that comprise agonists (Ia)
CC or antagonists (Ib) of human carcinoembryonal antigen (CEA). (Ia) are
CC used in vaccines to kill or inhibit carcinoma cells that express CEA or
CC its epitopes, particularly for treating gastrointestinal, breast,
CC pancreatic, bladder, ovarian, lung or prostatic carcinoma. They can also
CC be used to proliferate T cells, e.g. from vaccinated subjects, for use in
CC adoptive transfer therapy. (Ib) are used to inhibit CEA-specific immune
CC responses, e.g. in vaccinated subjects, to prevent an autoimmune reaction
CC to cancer immunotherapy (i.e. to prevent attack on normal but CEA-
CC expressing cells). (Ia) are more active than native sequence (I) and
CC generate a highly specific and systemic anti-CEA response. Cytotoxic T
CC cells generated recognize both (Ia) and native CEA epitopes. The present
CC sequence encodes a specifically claimed example of (Ia)
XX
SQ Sequence 27 BP; 5 A; 10 C; 6 G; 6 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	0.0945	Length:	27
Score:	43.00	Matches:	8
Percent Similarity:	100.00%	Conservative:	1
Best Local Similarity:	88.89%	Mismatches:	0
Query Match:	95.56%	Indels:	0
DB:	2	Gaps:	0

US-10-725-373-3 (1-9) x AAX56258 (1-27)

```
Qy          1 TyrLeuSerGlyAlaAspIleAsnLeu 9
              |||||||||||||||||:::||||||
Db          1 TACCTTTCGGGAGCGGACCTCAACCTC 27
```

RESULT 3
AAI72497
ID AAI72497 standard; DNA; 2105 BP.
XX
AC AAI72497;
XX
DT 16-MAY-2002 (first entry)
XX
DE CEA agonist coding sequence #2.
XX
KW CAP-1; epitope; carcinoembryonic antigen; CEA; agonist; immune response;
KW carcinoma; gastrointestinal; breast; pancreatic; bladder; ovarian; lung;
KW prostate; cancer; gene; ss.
XX
OS Synthetic.
XX

FH	Key	Location/Qualifiers
FT	CDS	1. .2105
FT		/*tag= a
FT		/product= "CEA agonist polypeptide"

XX
PN WO200210379-A2.
XX
PD 07-FEB-2002.
XX
PF 27-JUL-2001; 2001WO-CA001092.
XX
PR 31-JUL-2000; 2000US-0222043P.
XX
PA (AVET) AVENTIS PASTEUR LTD.
PA (THER-) THERION BIOLOGICS.
PA (USSH) US NAT CANCER INST.
XX
PI Berinstein N, Tartaglia J, Tine JA, Panicali DL, Gritz L;
PI Schlom J;
XX
DR WPI; 2002-206189/26.
XX
PT Carcinoembryonic antigen agonist polypeptide for inducing an immune
PT response in animal against antigen and for inhibiting an epitope antigen
PT expressing carcinoma cell, comprises a modified antigen epitope.
XX
PS Claim 4; Page 66-67; 69pp; English.

XX
CC This sequence encodes the carcinoembryonic antigen (CEA) agonist
CC polypeptide of the invention. This sequence represents the sequence given
CC in the Seq ID listing in the specification, and does not directly encodes
CC the CEA agonist polypeptide given in AAB47919. The CEA agonist contains a
CC modified CAP-1 epitope of CEA, in which position 6 is modified from Asp
CC to Asn to increase its immunogenicity. The CEA agonist polypeptide of the
CC invention, or DNA encoding it, are useful for: (i) inducing an immune
CC response in an animal directed against a CEA protein or fragment, CEA
CC agonist, a CEA epitope, a modified CEA epitope, cells expressing or
CC binding a CEA protein or fragment; and (ii) inhibiting a CEA epitope
CC expressing carcinoma cell, which is a gastrointestinal, breast,
CC pancreatic, bladder, ovarian, lung or prostate carcinoma cell in a
CC patient, hence is useful for manufacture of a medicament for the
CC treatment of cancer
XX
SQ Sequence 2105 BP; 555 A; 658 C; 441 G; 451 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	13.2	Length:	2105
Score:	43.00	Matches:	8
Percent Similarity:	100.00%	Conservative:	1
Best Local Similarity:	88.89%	Mismatches:	0
Query Match:	95.56%	Indels:	0
DB:	6	Gaps:	0

US-10-725-373-3 (1-9) x AAI72497 (1-2105)

```
Qy          1 TyrLeuSerGlyAlaAspIleAsnLeu 9
              ||||||||||||||||||:::|||||||
Db       1809 TACCTTTCGGGAGCGGACCTCAACCTC 1835
```

RESULT 4
AAH20121
ID AAH20121 standard; cDNA; 2106 BP.
XX
AC AAH20121;
XX
DT 08-AUG-2001 (first entry)
XX
DE Modified Carcinoembryonic antigen (CEA) encoding cDNA SEQ ID NO:111.
XX
KW Virus; adenovirus; poxvirus; alphavirus; immune response; gp100;
KW tumour antigen; CEA; carcinoembryonic antigen; immunostimulant;
KW cytostatic; immunotherapy; interferon-gamma; IFN-gamma; cancer; ss.
XX
OS Unidentified.
XX
PN WO200130382-A1.
XX
PD 03-MAY-2001.
XX
PF 20-OCT-2000; 2000WO-CA001253.
XX
PR 22-OCT-1999; 99US-0160879P.
PR 07-AUG-2000; 2000US-0223325P.
XX
PA (AVET) AVENTIS PASTEUR LTD.
XX
PI Berinstein N, Tartaglia J, Moingeon P, Barber B;
XX
DR WPI; 2001-308587/32.
DR P-PSDB; AAB97817.
XX
PT Inducing immune response to tumor antigen, useful in immunotherapy of
PT cancer, by administering the antigen to a lymphatic site.
XX
PS Disclosure; Fig 8; 60pp; English.
XX
CC The present invention describes a method for inducing an immune response,
CC in an animal, to a tumour antigen (Ag) comprising administering Ag, or
CC nucleic acid (I) that encodes it, to a lymphatic site. Cynomolgus monkeys
CC (Macaca fascicularis) were injected with a modified form of gp100 antigen
CC (a) into the left inguinal lymph node or (b) subcutaneously. Both animals
CC of (a) developed a cell-mediated response (indicated by production of
CC interferon-gamma from T lymphocytes when exposed to gp100 peptides), but
CC only 2 of 4 animals of (b) did so. Also animals in (a) produced a far
CC greater antibody response to gp100. The method is used in immunotherapy
CC of a wide range of cancers through induction of a specific immune
CC response (humoral and cellular) against the tumour antigens. When
CC administered to a lymphatic site, Ag (or (I)) induces a stronger immune
CC response than administration by other routes and may also break tolerance
CC to Ag. AAB97708 and AAB97709 represent gp100 epitopes; AAB97710 to
CC AAB97815 represent peptides derived from gp100 which stimulate interferon
CC (IFN)-gamma production; AAH20120 encodes the modified gp100 protein given
CC in AAB97816; AAH20121 encodes the modified carcinoembryonic antigen (CEA)
CC protein given in AAB97817; and AAB97818 represents a CEA modified antigen
CC peptide, all of which are used in the exemplification of the present
CC invention
XX
SQ Sequence 2106 BP; 559 A; 659 C; 442 G; 446 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	13.2	Length:	2106
Score:	43.00	Matches:	8
Percent Similarity:	100.00%	Conservative:	1
Best Local Similarity:	88.89%	Mismatches:	0
Query Match:	95.56%	Indels:	0
DB:	4	Gaps:	0

US-10-725-373-3 (1-9) x AAH20121 (1-2106)

```
Qy          1 TyrLeuSerGlyAlaAspIleAsnLeu 9
              ||||||||||||||||||:::|||||||
Db       1810 TACCTTTCGGGAGCGGACCTCAACCTC 1836
```

RESULT 5
AAD07347
ID AAD07347 standard; DNA; 2106 BP.
XX
AC AAD07347;
XX
DT 18-SEP-2001 (first entry)
XX
DE Modified carcinoembryonic antigen (CEA) DNA.
XX
KW Tumour-associated antigen; TAA; cytostatic; vaccine; gene therapy;
KW immune response; tetanus toxoid; TT; diphtheria toxoid; DT; prophylactic;
KW cancer; therapeutic; carcinoembryonic antigen; CEA; ds.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT CDS 1. .2106
FT /*tag= a
FT /product= "Modified carcinoembryonic antigen (CEA)"
XX
PN WO200149317-A2.
XX
PD 12-JUL-2001.
XX
PF 05-JAN-2001; 2001WO-CA000005.
XX
PR 05-JAN-2000; 2000US-0174587P.
XX
PA (AVET) AVENTIS PASTEUR LTD.
XX
PI Emtage P, Barber BH, Sambhara S, Sia CDY;
XX
DR WPI; 2001-441790/47.
DR P-PSDB; AAE05117.
XX
PT Enhancing immune response to antigen such as tumor antigen for treating
PT cancer in an animal involves administering an inducing agent to the
PT animal followed by administering inducing agent-antigen mixture.
XX

PS Claim 9; Fig 3; 62pp; English.
XX
CC The invention relates to a method of enhancing an immune response against
CC tumour-associated antigens (TAAs), such as GP100 and carcinoembryonic
CC antigen (CEA) in an animal. The method involves priming of the animal
CC with an inducing agent such as tetanus toxoid (TT) or diphtheria toxoid
CC (DT), subsequently followed by administration of an inducing agent-
CC antigen mixture. The method provides the enhancement or augmentation of
CC the immune response to the antigen and/or improves a vaccination protocol
CC by allowing use of less antigen. The immunisation of the animal with
CC tumour-associated antigen is useful for the prophylactic or therapeutic
CC treatment of cancer. The present DNA sequence encodes modified
CC carcinoembryonic antigen (CEA) related to the invention
XX
SQ Sequence 2106 BP; 559 A; 659 C; 442 G; 446 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	13.2	Length:	2106
Score:	43.00	Matches:	8
Percent Similarity:	100.00%	Conservative:	1
Best Local Similarity:	88.89%	Mismatches:	0
Query Match:	95.56%	Indels:	0
DB:	5	Gaps:	0

US-10-725-373-3 (1-9) x AAD07347 (1-2106)

```
Qy          1 TyrLeuSerGlyAlaAspIleAsnLeu 9
              |||||||||||||||||:::||||||
Db       1810 TACCTTTCGGGAGCGGACCTCAACCTC 1836
```

RESULT 6
AAI72489
ID AAI72489 standard; DNA; 2106 BP.
XX
AC AAI72489;
XX
DT 16-MAY-2002 (first entry)
XX
DE CEA agonist coding sequence #1.
XX
KW CAP-1; epitope; carcinoembryonic antigen; CEA; agonist; immune response;
KW carcinoma; gastrointestinal; breast; pancreatic; bladder; ovarian; lung;
KW prostate; cancer; gene; ds.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT CDS 1. .2106
FT /*tag= a
FT /product= "CEA agonist polypeptide"
XX
PN WO200210379-A2.
XX
PD 07-FEB-2002.
XX
PF 27-JUL-2001; 2001WO-CA001092.
XX
PR 31-JUL-2000; 2000US-0222043P.
XX
PA (AVET) AVENTIS PASTEUR LTD.
PA (THER-) THERION BIOLOGICS.
PA (USSH) US NAT CANCER INST.
XX
PI Berinstein N, Tartaglia J, Tine JA, Panicali DL, Gritz L;
PI Schlom J;
XX
DR WPI; 2002-206189/26.
DR P-PSDB; AAB47918.
XX
PT Carcinoembryonic antigen agonist polypeptide for inducing an immune
PT response in animal against antigen and for inhibiting an epitope antigen
PT expressing carcinoma cell, comprises a modified antigen epitope.
XX

PS Claim 4; Fig 1; 69pp; English.
XX
CC This sequence encodes the carcinoembryonic antigen (CEA) agonist
CC polypeptide of the invention. This sequence represents the sequence given
CC in the figures in the specification, and it directly encodes the CEA
CC agonist polypeptide given in AAB47918. The CEA agonist contains a
CC modified CAP-1 epitope of CEA, in which position 6 is modified from Asp
CC to Asn to increase its immunogenicity. The CEA agonist polypeptide of the
CC invention, or DNA encoding it, are useful for: (i) inducing an immune
CC response in an animal directed against a CEA protein or fragment, CEA
CC agonist, a CEA epitope, a modified CEA epitope, cells expressing or
CC binding a CEA protein or fragment; and (ii) inhibiting a CEA epitope
CC expressing carcinoma cell, which is a gastrointestinal, breast,
CC pancreatic, bladder, ovarian, lung or prostate carcinoma cell in a
CC patient, hence is useful for manufacture of a medicament for the
CC treatment of cancer
XX
SQ Sequence 2106 BP; 559 A; 658 C; 442 G; 447 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	13.2	Length:	2106
Score:	43.00	Matches:	8
Percent Similarity:	100.00%	Conservative:	1
Best Local Similarity:	88.89%	Mismatches:	0
Query Match:	95.56%	Indels:	0
DB:	6	Gaps:	0

US-10-725-373-3 (1-9) x AAI72489 (1-2106)

```
Qy          1 TyrLeuSerGlyAlaAspIleAsnLeu 9
              |||||||||||||||||:::||||||
Db       1810 TACCTTTCGGGAGCGGACCTCAACCTC 1836
```

RESULT 7
ADE13860
ID ADE13860 standard; DNA; 2106 BP.
XX
AC ADE13860;
XX
DT 29-JAN-2004 (first entry)
XX
DE CEA-CAP6D nucleotide sequence SEQ ID NO:23.
XX
KW carcinoembryonic antigen; CEA; CEA(6D)-1; 2; cytostatic; vaccine; cancer;
KW tumour antigen; immunotherapy; gene; ds.
XX
OS Unidentified.
XX
PN WO2003085087-A2.
XX
PD 16-OCT-2003.
XX
PF 09-APR-2003; 2003WO-US010916.
XX
PR 09-APR-2002; 2002US-0372972P.
XX
PA (AVET) AVENTIS PASTEUR LTD.
PA (THER-) THERION BIOLOGICS INC.
XX
PI Parrington M, Zhang L, Rovinski B, Gritz LR, Greenhalgh T;
XX
DR WPI; 2003-877029/81.
XX
PT New isolated DNA molecule comprising the carcinoembryonic antigen (6D)-
PT 1,2 sequence, useful for diagnosing, preventing and treating cancer, or
PT determining the effectiveness of a chemotherapeutic or other treatment
PT regimen.
XX
PS Example 1; SEQ ID NO 23; 56pp; English.
XX
CC The present invention describes an isolated DNA molecule comprising the
CC carcinoembryonic antigen (CEA) (6D)-1,2 sequence of 2106 bp (see
CC ADE13861), or its fragment. Also described: (1) an expression vector

```
CC   comprising the nucleic acid sequence CEA(6D)-1,2, or its fragment
CC   describes above; (2) a composition comprising the expression vector of
CC   (1) in a pharmaceutical carrier; and (3) preventing or treating cancer
CC   comprising administering to a host the expression vector of (1). CEA(6D)-
CC   1,2 has cytostatic activity, and can be used in vaccines. The CEA(6D)-1,2
CC   nucleic acid and target polypeptide are useful for diagnosing, preventing
CC   and treating cancer, predicting prognosis, or determining the
CC   effectiveness of a chemotherapeutic or other treatment regimen. The
CC   expression vector may be used for the insertion and expression of CEA(6D)
CC   -1,2 nucleic acid encoding tumour antigens for the immunotherapeutic
CC   treatment of cancer. The target polypeptides are useful in generating
CC   antibodies used in screening assays or for immunotherapy. The present
CC   sequence represents the CEA-CAP6D nucleotide sequence, which is given in
CC   comparison with CEA(6D)-1,2 in the exemplification of the present
CC   invention.
XX
SQ   Sequence 2106 BP; 559 A; 659 C; 442 G; 446 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:                13.2          Length:          2106
Score:                    43.00         Matches:         8
Percent Similarity:       100.00%       Conservative:    1
Best Local Similarity:    88.89%        Mismatches:      0
Query Match:              95.56%        Indels:          0
DB:                       10            Gaps:            0

US-10-725-373-3 (1-9) x ADE13860 (1-2106)

Qy          1 TyrLeuSerGlyAlaAspIleAsnLeu 9
              ||||||||||||||||||:::||||||
Db       1810 TACCTTTCGGGAGCGGACCTCAACCTC 1836

RESULT 8
ADE13861
ID   ADE13861 standard; DNA; 2106 BP.
XX
AC   ADE13861;
XX
DT   29-JAN-2004  (first entry)
XX
DE   CEA(6D)-1,2 nucleotide sequence SEQ ID NO:24.
XX
KW   carcinoembryonic antigen; CEA; CEA(6D)-1; 2; cytostatic; vaccine; cancer;
KW   tumour antigen; immunotherapy; gene; ds.
XX
OS   Unidentified.
XX
PN   WO2003085087-A2.
XX
PD   16-OCT-2003.
XX
PF   09-APR-2003; 2003WO-US010916.
XX
PR   09-APR-2002; 2002US-0372972P.
XX
PA   (AVET ) AVENTIS PASTEUR LTD.
PA   (THER-) THERION BIOLOGICS INC.
XX
PI   Parrington M,  Zhang L,  Rovinski B,  Gritz LR,  Greenhalgh T;
XX
DR   WPI; 2003-877029/81.
XX
PT   New isolated DNA molecule comprising the carcinoembryonic antigen (6D)-
PT   1,2 sequence, useful for diagnosing, preventing and treating cancer, or
PT   determining the effectiveness of a chemotherapeutic or other treatment
PT   regimen.
XX
PS   Claim 1; SEQ ID NO 24; 56pp; English.
XX
CC   The present invention describes an isolated DNA molecule comprising the
CC   carcinoembryonic antigen (CEA) (6D)-1,2 sequence of 2106 bp (see
CC   ADE13861), or its fragment. Also described: (1) an expression vector
CC   comprising the nucleic acid sequence CEA(6D)-1,2, or its fragment
CC   describes above; (2) a composition comprising the expression vector of
CC   (1) in a pharmaceutical carrier; and (3) preventing or treating cancer
CC   comprising administering to a host the expression vector of (1). CEA(6D)-
CC   1,2 has cytostatic activity, and can be used in vaccines. The CEA(6D)-1,2
CC   nucleic acid and target polypeptide are useful for diagnosing, preventing
CC   and treating cancer, predicting prognosis, or determining the
CC   effectiveness of a chemotherapeutic or other treatment regimen. The
CC   expression vector may be used for the insertion and expression of CEA(6D)
CC   -1,2 nucleic acid encoding tumour antigens for the immunotherapeutic
CC   treatment of cancer. The target polypeptides are useful in generating
CC   antibodies used in screening assays or for immunotherapy. The present
CC   sequence represents the CEA(6D)-1,2 nucleotide sequence, which is given
CC   in the exemplification of the present invention.
XX
SQ   Sequence 2106 BP; 574 A; 615 C; 435 G; 482 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:                13.2          Length:          2106
Score:                    43.00         Matches:         8
Percent Similarity:       100.00%       Conservative:    1
Best Local Similarity:    88.89%        Mismatches:      0
Query Match:              95.56%        Indels:          0
DB:                       10            Gaps:            0

US-10-725-373-3 (1-9) x ADE13861 (1-2106)

Qy          1 TyrLeuSerGlyAlaAspIleAsnLeu 9
              ||||||||||||||||||:::||||||
Db       1810 TACCTTTCGGGAGCGGACCTCAACCTC 1836

RESULT 9
AAI72490/c
ID   AAI72490 standard; DNA; 7958 BP.
XX
AC   AAI72490;
XX
DT   16-MAY-2002  (first entry)
XX
DE   H6-promoter human CEAmod/42K-promoted B7.1 insertion cassette.
XX
KW   CAP-1; epitope; carcinoembryonic antigen; CEA; agonist; immune response;
KW   carcinoma; gastrointestinal; breast; pancreatic; bladder; ovarian; lung;
KW   prostate; cancer; gene; ds.
XX
OS   Homo sapiens.
OS   Synthetic.
OS   Chimeric.
XX
FH   Key             Location/Qualifiers
FT   misc_feature    423. .827
FT                   /*tag=  a
FT                   /note= "ALVAC's C5 locus left flanking arm"
FT   CDS             complement(903. .3008)
FT                   /*tag=  b
FT                   /product= "CEA agonist peptide"
FT   promoter        complement(3009. .3132)
FT                   /*tag=  c
FT                   /label= Vaccinia_H6_promoter
FT   promoter        3210. .3275
FT                   /*tag=  d
FT                   /label= 42K_promoter
FT   CDS             3276. .4142
FT                   /*tag=  e
FT                   /product= "Human B7.1"
FT   misc_feature    4184. .5722
FT                   /*tag=  f
FT                   /note= "ALVAC's C5 locus right flanking arm#"
XX
PN   WO200210379-A2.
XX
PD   07-FEB-2002.
XX
PF   27-JUL-2001; 2001WO-CA001092.
```

XX
PR 31-JUL-2000; 2000US-0222043P.
XX
PA (AVET) AVENTIS PASTEUR LTD.
PA (THER-) THERION BIOLOGICS.
PA (USSH) US NAT CANCER INST.
XX
PI Berinstein N, Tartaglia J, Tine JA, Panicali DL, Gritz L;
PI Schlom J;
XX
DR WPI; 2002-206189/26.
XX
PT Carcinoembryonic antigen agonist polypeptide for inducing an immune
PT response in animal against antigen and for inhibiting an epitope antigen
PT expressing carcinoma cell, comprises a modified antigen epitope.
XX
PS Example 3; Fig 4; 69pp; English.
XX
CC This sequence represents ALVAC(2)-CEAmod/hB7.1. This is a coding sequence
CC containing the H6 promoted modified carcinoembryonic antigen (CEA)
CC agonist polypeptide of the invention. The CEA agonist contains a modified
CC CAP-1 epitope of CEA, in which position 6 is modified from Asp to Asn to
CC increase its immunogenicity. The CEA agonist polypeptide of the
CC invention, or DNA encoding it, are useful for: (i) inducing an immune
CC response in an animal directed against a CEA protein or fragment, CEA
CC agonist, a CEA epitope, a modified CEA epitope, cells expressing or
CC binding a CEA protein or fragment; and (ii) inhibiting a CEA epitope
CC expressing carcinoma cell, which is a gastrointestinal, breast,
CC pancreatic, bladder, ovarian, lung or prostate carcinoma cell in a
CC patient, hence is useful for manufacture of a medicament for the
CC treatment of cancer
XX
SQ Sequence 7958 BP; 2096 A; 1720 C; 1858 G; 2284 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	59.6	Length:	7958
Score:	43.00	Matches:	8
Percent Similarity:	100.00%	Conservative:	1
Best Local Similarity:	88.89%	Mismatches:	0
Query Match:	95.56%	Indels:	0
DB:	6	Gaps:	0

US-10-725-373-3 (1-9) x AAI72490 (1-7958)

```
Qy          1 TyrLeuSerGlyAlaAspIleAsnLeu 9
              |||||||||||||||||:::||||||
Db       1199 TACCTTTCGGGAGCGGACCTCAACCTC 1173
```

RESULT 10
ABD32542/c
ID ABD32542 standard; DNA; 108845 BP.
XX
AC ABD32542;
XX
DT 18-NOV-2004 (first entry)
XX
DE Mouse cancer-associated genomic DNA MD7-024.
XX
KW Mouse; ds; cancer-associated protein; gene; cytostatic; cancer;
KW leukaemia; lymphoma; CAP.
XX
OS Mus musculus.
XX
PN WO2004074320-A2.
XX
PD 02-SEP-2004.
XX
PF 17-FEB-2004; 2004WO-US004730.
XX
PR 14-FEB-2003; 2003US-00367094.
PR 14-MAR-2003; 2003US-00388838.
PR 15-APR-2003; 2003US-00417375.
PR 13-JUN-2003; 2003US-00461862.
PR 15-SEP-2003; 2003US-00663431.
PR 15-DEC-2003; 2003US-00737318.
XX
PA (SAGR-) SAGRES DISCOVERY INC.
XX
PI Morris DW, Morris DW, Malandro MS;
XX
DR WPI; 2004-652914/63.
XX
PT New isolated cancer-associated polynucleotides and polypeptides useful
PT for diagnosing, preventing or treating cancers, especially lymphoma and
PT leukemia, or in screening for agents that modulate cancer.
XX
PS disclosure; seqid 15; 310pp; English.
XX
CC The invention relates to an isolated nucleic acid comprising at least 10
CC contiguous nucleotides of any of the 233 polynucleotide sequences given
CC in the specification, or its complement. The nucleic acids encode cancer-
CC associated proteins. Also included are an expression vector comprising
CC the isolated nucleic acid cited above, a host cell comprising the above
CC recombinant nucleic acid or expression vector, a microarray for detecting
CC a cancer-associated (CA) nucleic acid comprising at least one probe
CC comprising at least 10 contiguous nucleotides of any of the above-
CC mentioned nucleotide sequences, an isolated polypeptide (encoded within
CC an open reading frame of a CA sequence selected from any of the 95
CC polynucleotide sequences as mentioned in the specification, or its
CC complement), an isolated antibody, (or its antigen binding fragment) that
CC binds to the above polypeptide, a hybridoma that produces the above
CC monoclonal antibody, a pharmaceutical composition comprising the above
CC antibody and a pharmaceutical excipient, a kit for detecting cancer
CC cells(comprising the antibody cited above, methods for diagnosing cancer
CC or for detecting the presence or absence of cancer cells in an
CC individual, a method for inhibiting growth of cancer cells in an
CC individual, a method for delivering a therapeutic agent to cancer cells
CC in an individual, an electronic library comprising the above
CC polynucleotide or polypeptide (or their fragments), methods of screening
CC for anticancer activity or for a bioactive agent capable of modulating
CC the activity of a CA protein (CAP), methods for detecting cancer
CC associated with expression of a polypeptide in a test cell sample, a
CC method for treating cancers and a method for inhibiting the expression of
CC CA gene in a cell. The composition and methods are useful for detecting,
CC diagnosing, preventing and treating cancers, especially lymphoma and
CC leukaemia. These may also be used in screening for agents that modulate
CC cancer. The present sequence is a mouse CAP genomic sequence. Note: The
CC sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 108845 BP; 31876 A; 21216 C; 21367 G; 30935 T; 0 U; 3451 Other;

Alignment Scores:

Pred. No.:	5.6e+03	Length:	108845
Score:	40.00	Matches:	7
Percent Similarity:	100.00%	Conservative:	2
Best Local Similarity:	77.78%	Mismatches:	0
Query Match:	88.89%	Indels:	0
DB:	13	Gaps:	0

US-10-725-373-3 (1-9) x ABD32542 (1-108845)

```
Qy          1 TyrLeuSerGlyAlaAspIleAsnLeu 9
              |||||||||||:::||||||||:::
Db     105618 TATTTGAGTGGGTCAGATATTAACATA 105592
```

RESULT 11
ABZ16916/c
ID ABZ16916 standard; DNA; 2000 BP.
XX
AC ABZ16916;
XX
DT 21-JAN-2003 (first entry)
XX
DE Arabidopsis thaliana stress regulated gene SEQ ID NO 4721.

XX
KW Arabidopsis thaliana; plant; gene; stress; transgenic; ds.
XX
OS Arabidopsis thaliana.
XX
PN WO200216655-A2.
XX
PD 28-FEB-2002.
XX
PF 24-AUG-2001; 2001WO-US026685.
XX
PR 24-AUG-2000; 2000US-0227866P.
PR 26-JAN-2001; 2001US-0264647P.
PR 22-JUN-2001; 2001US-0300111P.
XX
PA (SCRI) SCRIPPS RES INST.
PA (SYGN) SYNGENTA PARTICIPATIONS AG.
XX
PI Harper JF, Kreps J, Wang X, Zhu T;
XX
DR WPI; 2002-304127/34.
XX
PT Identifying a stress condition to which a plant cell has been exposed and
PT producing plants with increased tolerance to these abiotic stresses.
XX
PS Claim 144; SEQ ID NO 4721; 577pp + Sequence Listing; English.
XX
CC The invention relates to identifying a stress condition to which a plant
CC cell has been exposed, comprising: (a) contacting nucleic acid
CC representative of expressed polynucleotides in the plant cell with an
CC array or probes representative of the plant cell genome; and (b)
CC detecting a profile of expressed polynucleotides in the plant cell
CC characteristic of a stress response. The method is useful in the
CC production of transgenic plants, cells and seeds and in producing plants
CC with increased tolerance to abiotic stress. The present sequence is that
CC of an Arabidopsis thaliana stress regulated gene (ABZ12196-ABZ17574) used
CC in methods of the invention. Note: The sequence data for this patent is
CC not represented in the printed specification but is based on sequence
CC information supplied to Derwent by the European Patent Office
XX
SQ Sequence 2000 BP; 781 A; 288 C; 243 G; 688 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	102	Length:	2000
Score:	39.00	Matches:	7
Percent Similarity:	88.89%	Conservative:	1
Best Local Similarity:	77.78%	Mismatches:	1
Query Match:	86.67%	Indels:	0
DB:	6	Gaps:	0

US-10-725-373-3 (1-9) x ABZ16916 (1-2000)

```
Qy          1 TyrLeuSerGlyAlaAspIleAsnLeu 9
              ||||||||||||   |||||||||:::
Db        296 TATTTGAGTGGTGGCGATATTAACATG 270
```

RESULT 12
AAX56260
ID AAX56260 standard; DNA; 27 BP.
XX
AC AAX56260;
XX
DT 20-JUL-1999 (first entry)
XX
DE Carcinoembryonic antigen peptide agonist CAP-1 encoding DNA SEQ ID NO:6.
XX
KW Carcinoembryonic antigen; CEA; human; agonist; antagonist;
KW immune response; carcinoma; gastrointestinal; breast; pancreatic;
KW bladder; ovarian; lung; prostatic; T cell proliferation; cancer;
KW adoptive transfer therapy; autoimmune reaction; immunotherapy; ss.
XX
OS Homo sapiens.
OS Synthetic.
XX
PN WO9919478-A1.
XX
PD 22-APR-1999.
XX
PF 22-SEP-1998; 98WO-US019794.
XX
PR 10-OCT-1997; 97US-0061589P.
XX
PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
XX
PI Schlom J, Barzaga E, Zaremba S;
XX
DR WPI; 1999-326544/27.
XX
PT Peptide agonists and antagonists of carcinoembryonal antigen.
XX
PS Disclosure; Page 19; 72pp; English.
XX
CC The present invention describes peptides (A) that comprise agonists (Ia)
CC or antagonists (Ib) of human carcinoembryonal antigen (CEA). (Ia) are
CC used in vaccines to kill or inhibit carcinoma cells that express CEA or
CC its epitopes, particularly for treating gastrointestinal, breast,
CC pancreatic, bladder, ovarian, lung or prostatic carcinoma. They can also
CC be used to proliferate T cells, e.g. from vaccinated subjects, for use in
CC adoptive transfer therapy. (Ib) are used to inhibit CEA-specific immune
CC responses, e.g. in vaccinated subjects, to prevent an autoimmune reaction
CC to cancer immunotherapy (i.e. to prevent attack on normal but CEA-
CC expressing cells). (Ia) are more active than native sequence (I) and
CC generate a highly specific and systemic anti-CEA response. Cytotoxic T
CC cells generated recognize both (Ia) and native CEA epitopes. The present
CC sequence encodes a specifically claimed example of (Ia)
XX
SQ Sequence 27 BP; 6 A; 10 C; 5 G; 6 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	1.31	Length:	27
Score:	38.00	Matches:	7
Percent Similarity:	100.00%	Conservative:	2
Best Local Similarity:	77.78%	Mismatches:	0
Query Match:	84.44%	Indels:	0
DB:	2	Gaps:	0

US-10-725-373-3 (1-9) x AAX56260 (1-27)

```
Qy          1 TyrLeuSerGlyAlaAspIleAsnLeu 9
              |||||||||||||||:::::|||||
Db          1 TACCTTTCGGGAGCGAACCTCAACCTC 27
```

RESULT 13
ADL46174
ID ADL46174 standard; DNA; 30 BP.
XX
AC ADL46174;
XX
DT 17-JUN-2004 (first entry)
XX
DE Human CAP-1 tumour antigen fragment DNA, SEQ ID NO:8 #1.
XX
KW Tumour antigen; vaccine; immunoglobulin; CH3 fragment; fusion protein;
KW cancer; tumour; dendritic cell; endocytosis; immune response; cytostatic;
KW human; CAP-1; ds.
XX
OS Homo sapiens.
XX
PN WO2004024181-A1.
XX
PD 25-MAR-2004.
XX
PF 15-SEP-2003; 2003WO-CN000776.
XX
PR 13-SEP-2002; 2002CN-00136965.
XX

PA (LIJJ/) LI J.
XX
PI Li J;
XX
DR WPI; 2004-269898/25.
XX
PT Tumor-antigen vaccines with molecular weight far smaller than antigen-
PT antibody compound to enable endocytosis by dendritic cells to promote
PT very high immunoreaction for killing tumor cells.
XX
PS Example 2; SEQ ID NO 8; 28pp; Chinese.
XX
CC The invention relates to a tumour antigen vaccine comprising 7 or more
CC amino acids of a tumour antigen sequence joined to an immunoglobulin CH3
CC fragment. The invention also relates to DNA sequences encoding the
CC antigenic fusion polypeptide; expression vectors and host cells
CC comprising the DNA sequences; a process for recombinantly producing the
CC fusion polypeptide; and a vaccine composition comprising the fusion
CC polypeptide and a pharmaceutically acceptable carrier. The tumour antigen
CC used is preferably selected from 07-AP, AFP, ART-4, BAGE B, beta-
CC catenin/m, bcr-abl, CAMEL, CAP-1, CASP-8, CDC27m, CDK4/m, CEA, CT, Cyp-B,
CC DAM, ELF2M, ETV6-AML1, ETS, G250, GAGE, GnT-V, GP100, HAGE, HER- 2/NEU,
CC HLA-A*0201-R170I, HPV-E6, HPV-E7, EBNA, HSP70-2M, HST-2, hTERT, iCE,
CC KIAA0205, LAGE, LDLR/FUT, GDP-Lfucose, MAGE, MART-1/Melan-A, MCIR,
CC Myosin/m, MUC1, MUM-1,-2,-3, NA88-A, NY-ESO-1, P15, p190, P53, Pml/RAR
CC alpha, FRAME, PSA, PSM, RAGE, RAS, RU1, RU2, SAGE, SART-1, SART-3,
CC TEL/AML1, TPI/m, TRP-1, gp75, TRP-2, TRP-2/INT2 and WTI. The tumour
CC antigen vaccines of the invention are useful in cancer therapy. The
CC antigenic fusion protein used in the vaccine are much smaller than the
CC corresponding antibody-antigen complex, permitting them to be endocytosed
CC by dendritic cells and thereby resulting in a greatly increased anti-
CC tumour immune response. The present sequence represents DNA encoding a
CC fragment of the human CAP-1/CEA tumour antigen used in an example of the
CC invention. Note: The present sequence differs from that also referred to
CC as SEQ ID NO:8 () which is given on page 10 of the specification.
XX
SQ Sequence 30 BP; 6 A; 12 C; 5 G; 7 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	1.47	Length:	30
Score:	38.00	Matches:	7
Percent Similarity:	100.00%	Conservative:	2
Best Local Similarity:	77.78%	Mismatches:	0
Query Match:	84.44%	Indels:	0
DB:	12	Gaps:	0

US-10-725-373-3 (1-9) x ADL46174 (1-30)

```
Qy          1 TyrLeuSerGlyAlaAspIleAsnLeu 9
              |||||||||||||||:::::||||||
Db          1 TACCTTTCGGGAGCGAACCTCAACCTC 27
```

RESULT 14
ADL46175
ID ADL46175 standard; DNA; 64 BP.
XX
AC ADL46175;
XX
DT 17-JUN-2004 (first entry)
XX
DE Human immunoglobulin Fc fragment 5' PCR primer, SEQ ID NO:9 #1.
XX
KW Tumour antigen; vaccine; immunoglobulin; CH3 fragment; fusion protein;
KW cancer; tumour; dendritic cell; endocytosis; immune response; cytostatic;
KW human; Fc fragment; PCR; primer; ss.
XX
OS Homo sapiens.
XX
PN WO2004024181-A1.
XX
PD 25-MAR-2004.
XX
PF 15-SEP-2003; 2003WO-CN000776.
XX
PR 13-SEP-2002; 2002CN-00136965.
XX
PA (LIJJ/) LI J.
XX
PI Li J;
XX
DR WPI; 2004-269898/25.
XX
PT Tumor-antigen vaccines with molecular weight far smaller than antigen-
PT antibody compound to enable endocytosis by dendritic cells to promote
PT very high immunoreaction for killing tumor cells.
XX
PS Example 2; SEQ ID NO 9; 28pp; Chinese.
XX
CC The invention relates to a tumour antigen vaccine comprising 7 or more
CC amino acids of a tumour antigen sequence joined to an immunoglobulin CH3
CC fragment. The invention also relates to DNA sequences encoding the
CC antigenic fusion polypeptide; expression vectors and host cells
CC comprising the DNA sequences; a process for recombinantly producing the
CC fusion polypeptide; and a vaccine composition comprising the fusion
CC polypeptide and a pharmaceutically acceptable carrier. The tumour antigen
CC used is preferably selected from 07-AP, AFP, ART-4, BAGE B, beta-
CC catenin/m, bcr-abl, CAMEL, CAP-1, CASP-8, CDC27m, CDK4/m, CEA, CT, Cyp-B,
CC DAM, ELF2M, ETV6-AML1, ETS, G250, GAGE, GnT-V, GP100, HAGE, HER- 2/NEU,
CC HLA-A*0201-R170I, HPV-E6, HPV-E7, EBNA, HSP70-2M, HST-2, hTERT, iCE,
CC KIAA0205, LAGE, LDLR/FUT, GDP-Lfucose, MAGE, MART-1/Melan-A, MCIR,
CC Myosin/m, MUC1, MUM-1,-2,-3, NA88-A, NY-ESO-1, P15, p190, P53, Pml/RAR
CC alpha, FRAME, PSA, PSM, RAGE, RAS, RU1, RU2, SAGE, SART-1, SART-3,
CC TEL/AML1, TPI/m, TRP-1, gp75, TRP-2, TRP-2/INT2 and WTI. The tumour
CC antigen vaccines of the invention are useful in cancer therapy. The
CC antigenic fusion protein used in the vaccine are much smaller than the
CC corresponding antibody-antigen complex, permitting them to be endocytosed
CC by dendritic cells and thereby resulting in a greatly increased anti-
CC tumour immune response. Sequences ADL46175-ADL46176 represent PCR primers
CC used to amplify DNA encoding a human immunoglobulin Fc fragment in an
CC example of the invention. Note: The present sequence differs from that
CC also referred to as SEQ ID NO:9 () which is given on page 10 of the
CC specification.
XX
SQ Sequence 64 BP; 16 A; 22 C; 13 G; 13 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	3.47	Length:	64
Score:	38.00	Matches:	7
Percent Similarity:	100.00%	Conservative:	2
Best Local Similarity:	77.78%	Mismatches:	0
Query Match:	84.44%	Indels:	0
DB:	12	Gaps:	0

US-10-725-373-3 (1-9) x ADL46175 (1-64)

```
Qy          1 TyrLeuSerGlyAlaAspIleAsnLeu 9
              |||||||||||||||:::::||||||
Db         15 TACCTTTCGGGAGCGAACCTCAACCTC 41
```

RESULT 15
AAV57948/c
ID AAV57948 standard; DNA; 80 BP.
XX
AC AAV57948;
XX
DT 24-NOV-1998 (first entry)
XX
DE 708 vkcea primary reaction 1.2 oligonucleotide vkcea592r.
XX
KW Hepatitis B surface antigen; HBsAg; MHC class II-restricted peptide;
KW vaccination; vaccine; MHC class I molecule; immune response; cancer;
KW major histocompatibility complex molecule; pathogenic organism;
KW viral disease; autoimmune condition; allergy; PCR primer; ss.
XX
OS Synthetic.
XX

PN WO9833523-A1.
XX
PD 06-AUG-1998.
XX
PF 02-FEB-1998; 98WO-GB000325.
XX
PR 31-JAN-1997; 97GB-00001999.
PR 05-JUL-1997; 97GB-00014182.
PR 07-AUG-1997; 97GB-00016620.
PR 07-AUG-1997; 97GB-00016641.
PR 21-NOV-1997; 97GB-00024584.
XX
PA (BIOV-) BIOVATION LTD.
XX
PI Carr FJ, Carter G;
XX
DR WPI; 1998-437178/37.
XX
PT Immunogenic molecules - comprising nucleic acid and polypeptide portion,
PT from both of which peptide for presentation on major histocompatibility
PT complex molecules can be derived.
XX
PS Example 10; Page 60; 87pp; English.
XX
CC A molecule has been developed which comprises: (a) a nucleic acid portion
CC from which at least one peptide for presentation of MHC class I or class
CC II molecules, or both, may be derived, and (b) a polypeptide portion,
CC from which at least 1 peptide for presentation on MHC class I or class II
CC molecules, or both, may be derived. Also described in the present
CC invention is another molecule comprising: (a) a nucleic acid portion from
CC which at least 1 peptide for presentation on MHC class I or class II
CC molecules, or both, may be derived, and (b) a polypeptide portion
CC comprising a recognition domain capable of targeting the molecule to an
CC antigen presenting cell (APC), where the polypeptide portion does not
CC comprise a specific antigen binding site. The molecules can be used to
CC induce immune responses to treat or prevent, e.g. diseases caused by
CC pathogenic organisms, cancers, viral disease, e.g. HIV or hepatitis
CC infection, autoimmune conditions, e.g. Grave's disease, multiple
CC sclerosis, systemic lupus erythematosus, diabetes mellitus, Kawasaki's
CC disease, rheumatoid arthritis or allergies, e.g. atopic dermatitis,
CC allergic rhinitis, allergic conjunctivitis, atopic asthma or eczema. The
CC combination of DNA and polypeptide in the same molecule can give rise not
CC only to a combination of MHC class I- and MHC class II-mediated immune
CC responses but also to an enhancement of these responses compared to the
CC responses given by either DNA or polypeptide alone. The present sequence
CC represents an oligonucleotide used in an example from the present
CC invention
XX
SQ Sequence 80 BP; 16 A; 28 C; 19 G; 17 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	4.47	Length:	80
Score:	38.00	Matches:	7
Percent Similarity:	100.00%	Conservative:	2
Best Local Similarity:	77.78%	Mismatches:	0
Query Match:	84.44%	Indels:	0
DB:	2	Gaps:	0

US-10-725-373-3 (1-9) x AAV57948 (1-80)

```
Qy          1 TyrLeuSerGlyAlaAspIleAsnLeu 9
              |||||||||||||||::::::||||||
Db         68 TACCTGTCCGGCGCCAACCTGAACCTG 42
```

Search completed: May 17, 2005, 17:45:08
Job time : 333.5 secs

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: May 17, 2005, 06:13:14 ; Search time 11.25 Seconds
(without alignments)
76.973 Million cell updates/sec

Title: US-10-725-373-4
Perfect score: 45
Sequence: 1 YLSGANINL 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_79:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	43	95.6	702	2	A36319	carcinoembryonic a
2	37	82.2	506	2	C81704	monooxygenase-rela
3	36	80.0	554	2	A70904	probable acid-CoA
4	35	77.8	177	2	A30368	interleukin-1 rece
5	35	77.8	180	2	A39386	interleukin-1 rece
6	34	75.6	83	2	I64001	hypothetical prote
7	34	75.6	335	2	E71215	hypothetical prote
8	34	75.6	949	2	E75352	glycine cleavage s
9	34	75.6	954	2	AF2756	glycine cleavage s
10	34	75.6	954	2	E97537	glycine cleavage s
11	34	75.6	957	2	B85946	hypothetical prote
12	34	75.6	957	2	AC0873	glycine dehydrogen
13	34	75.6	957	2	F91100	glycine decarboxyl
14	34	75.6	957	2	S36834	glycine dehydrogen
15	34	75.6	959	2	AB0111	glycine dehydrogen
16	34	75.6	979	2	T16734	hypothetical prote
17	34	75.6	1003	2	A39521	glycine dehydrogen
18	34	75.6	1020	2	JN0124	glycine dehydrogen
19	34	75.6	1034	2	S50917	aminomethyltransfe
20	33	73.3	132	2	H86682	prophage pi1 prote
21	33	73.3	246	2	F84412	oxidoreductase [im
22	33	73.3	374	2	AF1296	phosphoribosylamin
23	33	73.3	374	2	AD1668	phosphoribosylamin
24	33	73.3	507	2	E71551	probable monooxyge
25	33	73.3	514	1	DWEBTT	threonine ammonia-
26	33	73.3	514	1	DWECTS	threonine ammonia-
27	33	73.3	514	2	AB0924	threonine ammonia-
28	33	73.3	514	2	B91217	threonine ammonia-
29	33	73.3	514	2	AG0474	threonine ammonia-
30	33	73.3	515	2	C86063	threonine ammonia-
31	33	73.3	741	2	A45771	2-5A-dependent RNA
32	33	73.3	797	2	B95377	hypothetical prote
33	33	73.3	950	2	D81821	glycine dehydrogen
34	33	73.3	958	2	E82994	glycine cleavage s
35	33	73.3	1286	2	S28634	adhesin AIDA-I pre
36	33	73.3	1321	2	E69129	protoporphyrin IX
37	33	73.3	1956	2	T16416	hypothetical prote
38	32	71.1	103	2	AE2556	hypothetical prote
39	32	71.1	236	2	T22220	hypothetical prote
40	32	71.1	248	2	T26461	hypothetical prote
41	32	71.1	259	2	S76576	hypothetical prote
42	32	71.1	262	2	C81384	shikimate 5-dehydr
43	32	71.1	272	2	T22562	hypothetical prote
44	32	71.1	336	2	B90071	ornithine transcar
45	32	71.1	343	2	T15192	hypothetical prote

ALIGNMENTS

RESULT 1
A36319
carcinoembryonic antigen precursor - human
N;Alternate names: CEA; meconium antigen 100
C;Species: Homo sapiens (man)
C;Date: 16-Sep-1992 #sequence_revision 16-Sep-1992 #text_change 09-Jul-2004
C;Accession: A36319; A27773; A31037; A25845; S08106; S31737; A44476; I54224; I59098; A26
R;Schrewe, H.; Thompson, J.; Bona, M.; Hefta, L.J.F.; Maruya, A.; Hassauer, M.; Shively,
Mol. Cell. Biol. 10, 2738-2748, 1990
A;Title: Cloning of the complete gene for carcinoembryonic antigen: analysis of its promo
A;Reference number: A36319; MUID:90258861; PMID:2342461
A;Accession: A36319
A;Molecule type: DNA
A;Residues: 1-702 <SCH>
A;Cross-references: UNIPROT:P06731; GB:M17303; NID:g178676; PIDN:AAB59513.1; PID:g178677
A;Note: the authors show the codons TTA for residue 641-Phe and CAG for residue 646-Thr
R;Beauchemin, N.; Benchimol, S.; Cournoyer, D.; Fuks, A.; Stanners, C.P.
Mol. Cell. Biol. 7, 3221-3230, 1987
A;Title: Isolation and characterization of full-length functional cDNA clones for human c
A;Reference number: A27773; MUID:88038876; PMID:3670312
A;Accession: A27773
A;Molecule type: mRNA
A;Residues: 1-702 <BEA>
A;Cross-references: GB:M29540; NID:g180222; PIDN:AAA51967.1; PID:g180223
R;Barnett, T.; Goebel, S.J.; Nothdurft, M.A.; Elting, J.J.
Genomics 3, 59-66, 1988
A;Title: Carcinoembryonic antigen family: characterization of cDNAs coding for NCA and CE
A;Reference number: A31037; MUID:89122014; PMID:3220478
A;Accession: A31037
A;Molecule type: mRNA
A;Residues: 1-702 <BAR>
A;Cross-references: GB:M29540; NID:g180222; PIDN:AAA51967.1; PID:g180223
A;Note: the authors translated the codon GTG for residue 130 as Leu
R;Oikawa, S.; Nakazato, H.; Kosaki, G.
Biochem. Biophys. Res. Commun. 142, 511-518, 1987
A;Title: Primary structure of human carcinoembryonic antigen (CEA) deduced from cDNA sequ
A;Reference number: A25845; MUID:87128144; PMID:3814146
A;Accession: A25845
A;Molecule type: mRNA
A;Residues: 5-702 <OIK>
A;Cross-references: GB:M15042; NID:g180198; PIDN:AAA51963.1; PID:g180199
R;Oikawa, S.
submitted to the EMBL Data Library, September 1989
A;Reference number: S08106
A;Accession: S08106
A;Molecule type: mRNA
A;Residues: 5-319,321-702 <OI2>
A;Cross-references: EMBL:X16455; NID:g29854; PIDN:CAA34474.1; PID:g825638
R;Barnett, T.
submitted to the EMBL Data Library, September 1991
A;Description: Genomic DNA sequence upstream of the translational start of the carcinoemb
A;Reference number: S31737

A;Accession: S31737
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-141 <BA2>
A;Cross-references: EMBL:X62151
R;Khan, W.N.; Fraengsmyr, L.; Teglund, S.; Israelsson, A.; Bremer, K.; Hammarstroem, S.
Genomics 14, 384-390, 1992
A;Title: Identification of three new genes and estimation of the size of the carcinoembr
A;Reference number: A44476; MUID:93052339; PMID:1427854
A;Accession: A44476
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: DNA
A;Residues: 35-141 <KHA>
R;Willcocks, T.C.; Craig, I.W.
Genomics 8, 492-500, 1990
A;Title: Characterization of the genomic organization of human carcinoembryonic antigen
A;Reference number: I54224; MUID:91139118; PMID:2286372
A;Accession: I54224
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-37 <RES>
A;Cross-references: GB:M60964; NID:g180215; PIDN:AAA51964.1; PID:g180217
R;Zimmermann, W.; Ortlieb, B.; Friedrich, R.; von Kleist, S.
Proc. Natl. Acad. Sci. U.S.A. 84, 2960-2964, 1987
A;Title: Isolation and characterization of cDNA clones encoding the human carcinoembryon
A;Reference number: I59098; MUID:87204247; PMID:3033671
A;Accession: I59098
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 331-702 <RE2>
A;Cross-references: GB:M16234; NID:g180240; PIDN:AAA51972.1; PID:g180241
R;Siepen, D.; Paxton, R.J.; Neumaier, M.; Shively, J.E.; Wagener, C.
Biochem. Biophys. Res. Commun. 147, 212-218, 1987
A;Title: Carcinoembryonic antigen (CEA) and two crossreacting antigens of 165 KD and 105
A;Reference number: A26831; MUID:87326349; PMID:3632664
A;Accession: A26831
A;Molecule type: protein
A;Residues: 35-64 <SIE>
R;Thomas, P.; Toth, C.A.
Biochem. Biophys. Res. Commun. 170, 391-396, 1990
A;Title: Carcinoembryonic antigen binding to Kupffer cells is via a peptide located at t
A;Reference number: A35490; MUID:90321257; PMID:2372297
A;Accession: A35490
A;Molecule type: protein
A;Residues: 'X',140-151,'X',153,'X',155-156 <THO>
A;Note: this is the amino terminal end of a fragment shown to mediate uptake by Kupffer
C;Comment: This heavily glycosylated membane protein of unknown function is a widely use
C;Comment: This protein may be processed at its C-terminus. It is anchored to the membra
C;Genetics:
A;Gene: GDB:CEA
A;Cross-references: GDB:119054; OMIM:114890
A;Map position: 19q13.2-19q13.2
A;Introns: 22/1; 142/1; 235/1; 320/1; 413/1; 498/1; 591/1; 676/1
C;Superfamily: carcinoembryonic antigen; carcinoembryonic antigen precursor amino-termin
C;Keywords: blocked carboxyl end; glycoprotein; lipoprotein; membrane protein; phosphati
F;1-138/Domain: carcinoembryonic antigen precursor amino-terminal homology <CEAN>
F;1-34/Domain: signal sequence #status predicted <SIG>
F;35-678/Product: carcinoembryonic antigen #status predicted <MAT>
F;160-217/Domain: immunoglobulin homology <IMM1>
F;252-301/Domain: immunoglobulin homology <IMM2>
F;338-395/Domain: immunoglobulin homology <IMM3>
F;516-573/Domain: immunoglobulin homology <IMM4>
F;608-657/Domain: immunoglobulin homology <IMM5>
F;679-702/Domain: carboxyl-terminal propeptide #status predicted <CTP>
F;678/Modified site: GPI-anchor ethanolamine amidated carboxyl end (Gly) (in mature form

  Query Match 95.6%; Score 43; DB 2; Length 702;
  Best Local Similarity 88.9%; Pred. No. 0.85;
  Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 YLSGANINL 9
|||||:||
Db 605 YLSGANLNL 613

RESULT 2
C81704
monooxygenase-related protein TC0425 [imported] - Chlamydia muridarum (strain Nigg)
C;Species: Chlamydia muridarum, Chlamydia trachomatis MoPn
C;Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 09-Jul-2004
C;Accession: C81704
R;Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey, J
, C.; Dodson, R.; Gwinn, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzberg,
Nucleic Acids Res. 28, 1397-1406, 2000
A;Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39.
A;Reference number: A81500; MUID:20150255; PMID:10684935
A;Accession: C81704
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-506 <TET>
A;Cross-references: UNIPROT:Q9PKP0; GB:AE002309; GB:AE002160; NID:g7190464; PIDN:AAF39281
A;Experimental source: strain Nigg (MoPn)
C;Genetics:
A;Gene: TC0425

  Query Match 82.2%; Score 37; DB 2; Length 506;
  Best Local Similarity 66.7%; Pred. No. 11;
  Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 YLSGANINL 9
|||| |:|:
Db 290 YLSGVNLNI 298

RESULT 3
A70904
probable acid-CoA ligase (EC 6.2.1.-) - Mycobacterium tuberculosis (strain H37RV)
C;Species: Mycobacterium tuberculosis
C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 09-Jul-2004
C;Accession: A70904
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.;
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A;Authors: Sqares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome s
A;Reference number: A70500; MUID:98295987; PMID:9634230
A;Accession: A70904
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-554 <COL>
A;Cross-references: UNIPROT:O07411; GB:Z97050; GB:AL123456; NID:g3256008; PIDN:CAB09749.1
A;Experimental source: strain H37Rv
C;Genetics:
A;Gene: fadD5
C;Superfamily: 4-coumarate-CoA ligase; acetate-CoA ligase homology
C;Keywords: acid-thiol ligase
F;72-525/Domain: acetate-CoA ligase homology <ACL>

  Query Match 80.0%; Score 36; DB 2; Length 554;
  Best Local Similarity 87.5%; Pred. No. 19;
  Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 YLSGANIN 8
| ||||||
Db 217 YTSGANIN 224

RESULT 4
A30368
interleukin-1 receptor antagonist secreted form precursor - human
C;Species: Homo sapiens (man)
C;Date: 07-Jun-1990 #sequence_revision 07-Jun-1990 #text_change 09-Jul-2004
C;Accession: A40956; I37894; A30368; S08160; S08159; A37822
R;Eisenberg, S.P.; Brewer, M.T.; Verderber, E.; Heimdal, P.; Brandhuber, B.J.; Thompson,
Proc. Natl. Acad. Sci. U.S.A. 88, 5232-5236, 1991

A;Title: Interleukin 1 receptor antagonist is a member of the interleukin 1 gene family:
A;Reference number: A40956; MUID:91271363; PMID:1828896
A;Accession: A40956
A;Molecule type: DNA
A;Residues: 1-177 <EIS>
A;Cross-references: UNIPROT:P18510; GB:M63099; NID:g186385; PIDN:AAB41943.1; PID:g186386
R;Lennard, A.; Gorman, P.; Carrier, M.; Griffiths, S.; Scotney, H.; Sheer, D.; Solari, R
Cytokine 4, 83-89, 1992
A;Title: Cloning and chromosome mapping of the human interleukin-1 receptor antagonist g
A;Reference number: I37894; MUID:92338323; PMID:1385987
A;Accession: I37894
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-177 <LEN>
A;Cross-references: EMBL:X64532; NID:g33798; PIDN:CAA45832.1; PID:g33799
R;Carter, D.B.; Deibel Jr., M.R.; Dunn, C.J.; Tomich, C.S.C.; Laborde, A.L.; Slightom, J
J.G.; Sieu, L.C.; Hardee, M.M.; Zurcher-Neely, H.A.; Reardon, I.M.; Heinrikson, R.L.; Tr
Nature 344, 633-638, 1990
A;Title: Purification, cloning, expression and biological characterization of an interle
A;Reference number: A30368; MUID:90220867; PMID:2139180
A;Accession: A30368
A;Molecule type: mRNA
A;Residues: 1-177 <CAR>
A;Cross-references: GB:X53296; NID:g32578; PIDN:CAA37386.1; PID:g32579
A;Note: parts of this sequence, including the amino end of the mature protein, were conf
R;Eisenberg, S.P.; Evans, R.J.; Arend, W.P.; Verderber, E.; Brewer, M.T.; Hannum, C.H.;
Nature 343, 341-346, 1990
A;Title: Primary structure and functional expression from complementary DNA of a human i
A;Reference number: S08160; MUID:90136921; PMID:2137201
A;Accession: S08160
A;Status: not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-177 <EI2>
A;Cross-references: GB:X52015; NID:g32576; PIDN:CAA36262.1; PID:g32577
R;Hannum, C.H.; Wilcox, C.J.; Arend, W.P.; Joslin, F.G.; Dripps, D.J.; Heimdal, P.L.; Ar
Nature 343, 336-340, 1990
A;Title: Interleukin-1 receptor antagonist activity of a human interleukin-1 inhibitor.
A;Reference number: S08159; MUID:90136920; PMID:2137200
A;Accession: S08159
A;Molecule type: protein
A;Residues: 26-75;97-108;110-116;120-131;163-176 <HAN>
R;Bienkowski, M.J.; Eessalu, T.E.; Berger, A.E.; Truesdell, S.E.; Shelly, J.A.; Laborde,
J. Biol. Chem. 265, 14505-14511, 1990
A;Title: Purification and characterization of interleukin 1 receptor level antagonist pr
A;Reference number: A37822; MUID:90354444; PMID:2143761
A;Accession: A37822
A;Molecule type: protein
A;Residues: 26-52;70-77;122-127;170-175 <BIE>
A;Experimental source: culture medium, PMA-stimulated THP-1 cells
C;Comment: For an alternative splice form, see PIR:A39386
C;Genetics:
A;Gene: GDB:IL1RN
A;Cross-references: GDB:125897; OMIM:147679
A;Map position: 2q14.2-2q14.2
A;Introns: 39/2; 69/1; 106/3
C;Superfamily: interleukin-1
C;Keywords: alternative splicing; cytokine receptor; extracellular protein; glycoprotein
F;1-25/Domain: signal sequence #status predicted <SIG>
F;26-177/Product: interleukin-1 receptor antagonist #status experimental <MAT>
F;109/Binding site: carbohydrate (Asn) (covalent) #status experimental

Query Match 77.8%; Score 35; DB 2; Length 177;
Best Local Similarity 66.7%; Pred. No. 9;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 YLSGANINL 9
|| | |:||
Db 59 YLQGPNVNL 67

RESULT 5
A39386
interleukin-1 receptor antagonist, long intracellular splice form - human
N;Contains: interleukin-1 receptor antagonist, short intracellular splice form
C;Species: Homo sapiens (man)
C;Date: 28-Feb-1992 #sequence_revision 11-Apr-1997 #text_change 09-Jul-2004
C;Accession: I37893; A39386
R;Muzio, M.; Polentarutti, N.; Sironi, M.; Poli, G.; De Gioia, L.; Introna, M.; Mantovani
J. Exp. Med. 182, 623-628, 1995
A;Title: Cloning and characterization of a new isoform of the interleukin 1 receptor anta
A;Reference number: I37893; MUID:95355865; PMID:7629520
A;Accession: I37893
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-180 <RES>
A;Cross-references: UNIPROT:P18510; EMBL:X84348; NID:g1008970; PIDN:CAA59087.1; PID:g100
R;Haskill, S.; Martin, G.; Van Le, L.; Morris, J.; Peace, A.; Bigler, C.F.; Jaffe, G.J.;
Proc. Natl. Acad. Sci. U.S.A. 88, 3681-3685, 1991
A;Title: cDNA cloning of an intracellular form of the human interleukin 1 receptor antago
A;Reference number: A39386; MUID:91219436; PMID:1827201
A;Accession: A39386
A;Molecule type: mRNA
A;Residues: 1-3,25-180 <HAS>
A;Cross-references: GB:M55646; NID:g186291; PIDN:AAA59138.1; PID:g186292
C;Comment: For an alternative splice form, see PIR:A30368
C;Genetics:
A;Gene: GDB:IL1RN
A;Cross-references: GDB:125897; OMIM:147679
A;Map position: 2q14.2-2q14.2
C;Superfamily: interleukin-1
C;Keywords: alternative splicing; cytokine receptor
F;1-180/Product: interleukin-1 receptor antagonist, long intracellular splice form #statu
F;1-3,25-180/Product: interleukin-1 receptor antagonist, short intracellular splice form

Query Match 77.8%; Score 35; DB 2; Length 180;
Best Local Similarity 66.7%; Pred. No. 9.2;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 YLSGANINL 9
|| | |:||
Db 62 YLQGPNVNL 70

RESULT 6
I64001
hypothetical protein HI0117 - Haemophilus influenzae (strain Rd KW20)
C;Species: Haemophilus influenzae
C;Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 31-Oct-1997
C;Accession: I64001
R;Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage, A.
; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman, J.
, D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghagen, N.S.M.
Science 269, 496-512, 1995
A;Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter,
A;Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.
A;Reference number: A64000; MUID:95350630; PMID:7542800
A;Accession: I64001
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-83 <TIGR>
A;Cross-references: GB:U32697; GB:L42023; NID:g1573067; PID:g1573069; TIGR:HI0117

Query Match 75.6%; Score 34; DB 2; Length 83;
Best Local Similarity 100.0%; Pred. No. 6.4;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 LSGANIN 8
|||||||
Db 9 LSGANIN 15

RESULT 7
E71215
hypothetical protein PH1989 - Pyrococcus horikoshii
C;Species: Pyrococcus horikoshii
C;Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 12-Jul-2004

C;Accession: E71215
R;Kawarabayasi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; Sekir
M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguchi
DNA Res. 5, 55-76, 1998
A;Title: Complete sequence and gene organization of the genome of a hyper-thermophilic a
A;Reference number: A71000; MUID:98344137; PMID:9679194
A;Accession: E71215
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-335 <KAW>
A;Cross-references: UNIPROT:O57713; GB:AP000007; NID:g3236134; PIDN:BAA31116.1; PID:g325
A;Experimental source: strain OT3
A;Note: this accession replaces an interim accession for a sequence replaced by GenBank
C;Genetics:
A;Gene: PH1989

Query Match 75.6%; Score 34; DB 2; Length 335;
Best Local Similarity 66.7%; Pred. No. 29;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 YLSGANINL 9
:| |||||:
Db 59 FLKGANINV 67

RESULT 8
E75352
glycine cleavage system P protein - Deinococcus radiodurans (strain R1)
C;Species: Deinococcus radiodurans
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jul-2004
C;Accession: E75352
R;White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;
, M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A;Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A;Reference number: A75250; MUID:20036896; PMID:10567266
A;Accession: E75352
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-949 <WHI>
A;Cross-references: UNIPROT:Q9RTF5; GB:AE002021; GB:AE000513; NID:g6459573; PIDN:AAF1136
A;Experimental source: strain R1
C;Genetics:
A;Gene: DR1809
A;Map position: 1

Query Match 75.6%; Score 34; DB 2; Length 949;
Best Local Similarity 75.0%; Pred. No. 90;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 YLSGANIN 8
|| |||:|
Db 668 YLDGANMN 675

RESULT 9
AF2756
glycine cleavage system protein P2 gcvP [imported] - Agrobacterium tumefaciens (strain C
C;Species: Agrobacterium tumefaciens
C;Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 09-Jul-2004
C;Accession: AF2756
R;Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, L
erage, G.; Gillet, W.; Grant, C.; Guenthner, D.; Kutyavin, T.; Levy, R.; Li, M.; McClell
; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A;Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
ster, E.W.
A;Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A;Reference number: AB2577; MUID:21608550; PMID:11743193
A;Accession: AF2756
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-954 <KUR>
A;Cross-references: UNIPROT:Q8UFD6; GB:AE008688; PIDN:AAL42468.1; PID:g17739884; GSPDB:GN
A;Experimental source: strain C58 (Dupont)
C;Genetics:
A;Gene: gcvP
A;Map position: circular chromosome

Query Match 75.6%; Score 34; DB 2; Length 954;
Best Local Similarity 75.0%; Pred. No. 91;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 YLSGANIN 8
|| |||:|
Db 675 YLDGANMN 682

RESULT 10
E97537
glycine cleavage system protein P2 (PA2445) [imported] - Agrobacterium tumefaciens (strai
C;Species: Agrobacterium tumefaciens
C;Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 09-Jul-2004
C;Accession: E97537
R;Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman,
A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.;
Science 294, 2323-2328, 2001
A;Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tume
A;Reference number: A97359; MUID:21608551; PMID:11743194
A;Accession: E97537
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-954 <KUR>
A;Cross-references: UNIPROT:Q8UFD6; GB:AE007869; PIDN:AAK87254.1; PID:g15156542; GSPDB:GN
C;Genetics:
A;Gene: AGR_C_2699
A;Map position: circular chromosome

Query Match 75.6%; Score 34; DB 2; Length 954;
Best Local Similarity 75.0%; Pred. No. 91;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 YLSGANIN 8
|| |||:|
Db 675 YLDGANMN 682

RESULT 11
B85946
hypothetical protein gcvP [imported] - Escherichia coli (strain O157:H7, substrain EDL933
C;Species: Escherichia coli
C;Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C;Accession: B85946
R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew,
iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
Nature 409, 529-533, 2001
A;Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A;Reference number: A85480; MUID:21074935; PMID:11206551
A;Accession: B85946
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-957 <STO>
A;Cross-references: GB:AE005174; NID:g12517430; PIDN:AAG58030.1; GSPDB:GN00145; UWGP:Z424
A;Experimental source: strain O157:H7, substrain EDL933
C;Genetics:
A;Gene: gcvP

Query Match 75.6%; Score 34; DB 2; Length 957;
Best Local Similarity 75.0%; Pred. No. 91;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 YLSGANIN 8
|| |||:|
Db 679 YLDGANMN 686

RESULT 12
AC0873
glycine dehydrogenase (decarboxylating) [imported] - Salmonella enterica subsp. enterica
C;Species: Salmonella enterica subsp. enterica serovar Typhi
A;Note: this species has also been called Salmonella typhi
C;Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002
C;Accession: AC0873
R;Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,
th, T.; Connerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,
, S.; Moule, S.; O'Gaora, P.
Nature 413, 848-852, 2001
A;Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;
A;Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov
A;Reference number: AB0502; MUID:21534947; PMID:11677608
A;Accession: AC0873
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-957 <PAR>
A;Cross-references: GB:AL513382; PIDN:CAD02883.1; PID:g16504136; GSPDB:GN00176
C;Genetics:
A;Gene: STY3209

Query Match 75.6%; Score 34; DB 2; Length 957;
Best Local Similarity 75.0%; Pred. No. 91;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 YLSGANIN 8
|| |||:|
Db 679 YLDGANMN 686

RESULT 13
F91100
glycine decarboxylase [imported] - Escherichia coli (strain O157:H7, substrain RIMD 0509
C;Species: Escherichia coli
C;Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 09-Jul-2004
C;Accession: F91100
R;Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A;Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genc
A;Reference number: A99629; MUID:21156231; PMID:11258796
A;Accession: F91100
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-957 <HAY>
A;Cross-references: UNIPROT:Q8XD33; GB:BA000007; PIDN:BAB37197.1; PID:g13363246; GSPDB:G
A;Experimental source: strain O157:H7, substrain RIMD 0509952
C;Genetics:
A;Gene: ECs3774

Query Match 75.6%; Score 34; DB 2; Length 957;
Best Local Similarity 75.0%; Pred. No. 91;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 YLSGANIN 8
|| |||:|
Db 679 YLDGANMN 686

RESULT 14
S36834
glycine dehydrogenase (decarboxylating) (EC 1.4.4.2) - Escherichia coli (strain K-12)
C;Species: Escherichia coli
C;Date: 22-Jan-1994 #sequence_revision 10-Nov-1995 #text_change 09-Jul-2004
C;Accession: S36834; I41232; G65074
R;Okamura-Ikeda, K.; Ohmura, Y.; Fujiwara, K.; Motokawa, Y.
Eur. J. Biochem. 216, 539-548, 1993
A;Title: Cloning and nucleotide sequence of the gcv operon encoding the Escherichia coli
A;Reference number: S36832; MUID:93387305; PMID:8375392
A;Accession: S36834
A;Molecule type: DNA
A;Residues: 1-957 <OKA>
A;Cross-references: UNIPROT:P33195; EMBL:X73958; NID:g403342; PIDN:CAA52146.1; PID:g40334
R;Stauffer, L.T.; Fogarty, S.J.; Stauffer, G.V.
Gene 142, 17-22, 1994
A;Title: Characterization of the Escherichia coli gcv operon.
A;Reference number: I41231; MUID:94237484; PMID:8181752
A;Accession: I41232
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-957 <RES>
A;Cross-references: GB:L20872; NID:g304890; PIDN:AAA23867.1; PID:g304892
R;Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Col
.A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A;Title: The complete genome sequence of Escherichia coli K-12.
A;Reference number: A64720; MUID:97426617; PMID:9278503
A;Accession: G65074
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-957 <BLAT>
A;Cross-references: GB:AE000373; GB:U00096; NID:g2367173; PIDN:AAC75941.1; PID:g1789269;
A;Experimental source: strain K-12, substrain MG1655
C;Genetics:
A;Gene: gcvP; gcvHP
C;Keywords: oxidoreductase; phosphoprotein; pyridoxal phosphate
F;708/Binding site: pyridoxal phosphate (Lys) (covalent) #status predicted

Query Match 75.6%; Score 34; DB 2; Length 957;
Best Local Similarity 75.0%; Pred. No. 91;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 YLSGANIN 8
|| |||:|
Db 679 YLDGANMN 686

RESULT 15
AB0111
glycine dehydrogenase (decarboxylating) (EC 1.4.4.2) [imported] - Yersinia pestis (strain
C;Species: Yersinia pestis
C;Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 09-Jul-2004
C;Accession: AB0111
R;Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.;
deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;
il, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrell,
Nature 413, 523-527, 2001
A;Title: Genome sequence of Yersinia pestis, the causative agent of plague.
A;Reference number: AB0001; MUID:21470413; PMID:11586360
A;Accession: AB0111
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-959 <KUR>
A;Cross-references: UNIPROT:Q8ZHI8; GB:AL590842; PIDN:CAC89749.1; PID:g15978976; GSPDB:GN
C;Genetics:
A;Gene: gcvP
C;Keywords: oxidoreductase

Query Match 75.6%; Score 34; DB 2; Length 959;
Best Local Similarity 75.0%; Pred. No. 91;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 YLSGANIN 8
|| |||:|
Db 679 YLDGANMN 686

Search completed: May 17, 2005, 06:20:03
Job time : 12.25 secs

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: May 17, 2005, 06:13:14 ; Search time 51.75 Seconds
(without alignments)
89.057 Million cell updates/sec

Title: US-10-725-373-4
Perfect score: 45
Sequence: 1 YLSGANINL 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : UniProt_03:*
1: uniprot_sprot:*
2: uniprot_trembl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	43	95.6	420	2	Q68DM9	Q68dm9 homo sapien
2	43	95.6	702	1	CEA5_HUMAN	P06731 homo sapien
3	43	95.6	702	2	Q8N4D0	Q8n4d0 homo sapien
4	37	82.2	261	2	Q6AJS0	Q6ajs0 desulfotale
5	37	82.2	265	2	Q73XB6	Q73xb6 mycobacteri
6	37	82.2	506	2	Q9PKP0	Q9pkp0 chlamydia m
7	36	80.0	131	2	Q35903	Q35903 strongyloce
8	36	80.0	253	2	Q85I11	Q85i11 strongyloce
9	36	80.0	281	2	Q81CR9	Q81cr9 bacillus ce
10	36	80.0	518	2	Q7DAC6	Q7dac6 mycobacteri
11	36	80.0	554	2	O07411	O07411 mycobacteri
12	36	80.0	554	2	Q7U2P2	Q7u2p2 mycobacteri
13	36	80.0	824	2	Q6BHN8	Q6bhn8 debaryomyce
14	35	77.8	159	2	Q7RTZ4	Q7rtz4 homo sapien
15	35	77.8	177	1	IL1X_HUMAN	P18510 homo sapien
16	35	77.8	177	2	Q866R8	Q866r8 macaca fasc
17	35	77.8	315	2	Q7PUJ7	Q7puj7 anopheles g
18	35	77.8	318	2	Q8GH85	Q8gh85 bacillus th
19	35	77.8	448	2	Q88XM9	Q88xm9 lactobacill
20	35	77.8	846	2	Q9AIP5	Q9aip5 carsonella
21	35	77.8	955	1	GCSP_BRAJA	Q89i86 bradyrhizob
22	35	77.8	959	1	GCSP_SYNPX	Q7u3q5 synechococc
23	35	77.8	962	1	GCSP_PROMM	Q7v411 prochloroco
24	35	77.8	964	1	GCSP_PROMA	Q7v9k4 prochloroco
25	35	77.8	969	2	Q7UZJ5	Q7uzj5 prochloroco
26	35	77.8	990	2	Q6N344	Q6n344 rhodopseudo
27	35	77.8	1041	2	Q68ST1	Q68st1 pleurotus d
28	35	77.8	1611	2	Q8RFV3	Q8rfv3 fusobacteri
29	35	77.8	1619	2	Q7P2Z7	Q7p2z7 fusobacteri
30	35	77.8	1630	2	Q7Q073	Q7q073 anopheles g
31	34	75.6	189	2	Q7R4D9	Q7r4d9 giardia lam
32	34	75.6	214	2	Q9LDX7	Q9ldx7 oryza sativ
33	34	75.6	240	2	Q6SG72	Q6sg72 uncultured
34	34	75.6	240	2	Q6UCY1	Q6ucy1 uncultured
35	34	75.6	240	2	Q7BKE9	Q7bke9 gamma-prote
36	34	75.6	292	2	Q8WU56	Q8wu56 homo sapien
37	34	75.6	293	2	Q7VB23	Q7vb23 prochloroco
38	34	75.6	335	1	YJ89_PYRHO	O57713 pyrococcus
39	34	75.6	385	2	P94599	P94599 bacillus th
40	34	75.6	395	2	Q9LU43	Q9lu43 arabidopsis
41	34	75.6	444	2	Q86LS6	Q86ls6 caenorhabdi
42	34	75.6	500	1	GCSB_RHOBA	Q7unh1 rhodopirell
43	34	75.6	500	2	Q73HP6	Q73hp6 wolbachia p
44	34	75.6	514	2	Q7VRL9	Q7vrl9 candidatus
45	34	75.6	561	2	Q8D1V9	Q8d1v9 wiggleswort

ALIGNMENTS

RESULT 1
Q68DM9
ID Q68DM9 PRELIMINARY; PRT; 420 AA.
AC Q68DM9;
DT 25-OCT-2004 (TrEMBLrel. 28, Created)
DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE Hypothetical protein DKFZp781M2392.
GN Name=DKFZp781M2392;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Colon carcinoma;
RG The German cDNA Consortium;
RA Poustka A., Albert R., Moosmayer P., Schupp I., Wellenreuther R.,
RA Mewes H.W., Weil B., Amid C., Osanger A., Fobo G., Han M., Wiemann S.;
RL Submitted (AUG-2004) to the EMBL/GenBank/DDBJ databases.
DR EMBL; CR749337; CAH18191.1; -.
DR InterPro; IPR001589; Actbind_actnin.
DR InterPro; IPR003599; Ig.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003598; Ig_c2.
DR Pfam; PF00047; ig; 3.
DR SMART; SM00409; IG; 3.
DR SMART; SM00408; IGc2; 3.
DR PROSITE; PS00019; ACTININ_1; UNKNOWN_1.
DR PROSITE; PS50835; IG_LIKE; 3.
KW Hypothetical protein.
SQ SEQUENCE 420 AA; 45508 MW; 6E30C0B4A00D0F59 CRC64;

Query Match 95.6%; Score 43; DB 2; Length 420;
Best Local Similarity 88.9%; Pred. No. 3.9;
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 YLSGANINL 9
              |||||:||
Db        323 YLSGANLNL 331
```

RESULT 2
CEA5_HUMAN
ID CEA5_HUMAN STANDARD; PRT; 702 AA.
AC P06731;
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 25-OCT-2004 (Rel. 45, Last annotation update)
DE Carcinoembryonic antigen-related cell adhesion molecule 5 precursor
DE (Carcinoembryonic antigen) (CEA) (Meconium antigen 100) (CD66e
DE antigen).
GN Name=CEACAM5; Synonyms=CEA;
OS Homo sapiens (Human).

```
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=90258861; PubMed=2342461;
RA   Schrewe H., Thompson J., Bona M., Hefta L.J.F., Maruya A.,
RA   Hassauer M., Shively J.E., von Kleist S., Zimmermann W.;
RT   "Cloning of the complete gene for carcinoembryonic antigen: analysis
RT   of its promoter indicates a region conveying cell type-specific
RT   expression.";
RL   Mol. Cell. Biol. 10:2738-2748(1990).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=88038876; PubMed=3670312;
RA   Beauchemin N., Benchimol S., Cournoyer D., Fuks A., Stanners C.P.;
RT   "Isolation and characterization of full-length functional cDNA clones
RT   for human carcinoembryonic antigen.";
RL   Mol. Cell. Biol. 7:3221-3230(1987).
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=89122014; PubMed=3220478;
RA   Barnett T., Goebel S.J., Nothdurft M.A., Elting J.J.;
RT   "Carcinoembryonic antigen family: characterization of cDNAs coding for
RT   NCA and CEA and suggestion of nonrandom sequence variation in their
RT   conserved loop-domains.";
RL   Genomics 3:59-66(1988).
RN   [4]
RP   SEQUENCE OF 5-702 FROM N.A.
RX   MEDLINE=87128144; PubMed=3814146;
RA   Oikawa S., Nakazato H., Kosaki G.;
RT   "Primary structure of human carcinoembryonic antigen (CEA) deduced
RT   from cDNA sequence.";
RL   Biochem. Biophys. Res. Commun. 142:511-518(1987).
RN   [5]
RP   SEQUENCE OF 331-702 FROM N.A.
RX   MEDLINE=87204247; PubMed=3033671;
RA   Zimmermann W., Ortlieb B., Friedrich R., von Kleist S.;
RT   "Isolation and characterization of cDNA clones encoding the human
RT   carcinoembryonic antigen reveal a highly conserved repeating
RT   structure.";
RL   Proc. Natl. Acad. Sci. U.S.A. 84:2960-2964(1987).
CC   -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.
CC   -!- TISSUE SPECIFICITY: Found in adenocarcinomas of endodermally
CC       derived digestive system epithelium and fetal colon.
CC   -!- PTM: Complex immunoreactive glycoprotein with a MW of 180 kDa
CC       comprising 60% carbohydrate.
CC   -!- SIMILARITY: Belongs to the immunoglobulin superfamily. CEA family.
CC   -!- SIMILARITY: Contains 7 immunoglobulin-like domains.
CC   -!- DATABASE: NAME=PROW; NOTE=CD guide CD66e entry;
CC       WWW="http://www.ncbi.nlm.nih.gov/prow/cd/cd66e.htm".
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M17303; AAB59513.1; -.
DR   EMBL; M59262; AAA62835.1; ALT_SEQ.
DR   EMBL; M59255; AAA62835.1; JOINED.
DR   EMBL; M59256; AAA62835.1; JOINED.
DR   EMBL; M59257; AAA62835.1; JOINED.
DR   EMBL; M59258; AAA62835.1; JOINED.
DR   EMBL; M59259; AAA62835.1; JOINED.
DR   EMBL; M59260; AAA62835.1; JOINED.
DR   EMBL; M59261; AAA62835.1; JOINED.
DR   EMBL; M59709; -; NOT_ANNOTATED_CDS.
DR   EMBL; M59710; -; NOT_ANNOTATED_CDS.
DR   EMBL; M29540; AAA51967.1; -.
DR   EMBL; X16455; CAA34474.1; -.
DR   EMBL; M15042; AAA51963.1; -.
DR   EMBL; M16234; AAA51972.1; -.
DR   PIR; A36319; A36319.
DR   PDB; 1E07; Model; A=35-676.
DR   Genew; HGNC:1817; CEACAM5.
DR   MIM; 114890; -.
DR   GO; GO:0005887; C:integral to plasma membrane; TAS.
DR   InterPro; IPR007110; Ig-like.
DR   Pfam; PF00047; ig; 6.
DR   PROSITE; PS50835; IG_LIKE; 6.
KW   3D-structure; Glycoprotein; GPI-anchor; Immunoglobulin domain;
KW   Lipoprotein; Membrane; Repeat; Signal.
FT   SIGNAL        1     34
FT   CHAIN        35    685       Carcinoembryonic antigen-related cell
FT                                 adhesion molecule 5.
FT   PROPEP      686    702       Removed in mature form (Potential).
FT   DOMAIN       35    144       Ig-like 1.
FT   DOMAIN      146    237       Ig-like 2.
FT   DOMAIN      238    322       Ig-like 3.
FT   DOMAIN      324    415       Ig-like 4.
FT   DOMAIN      416    498       Ig-like 5.
FT   DOMAIN      502    593       Ig-like 6.
FT   DOMAIN      594    677       Ig-like 7.
FT   CARBOHYD    104    104       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    115    115       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    152    152       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    182    182       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    197    197       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    204    204       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    208    208       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    246    246       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    256    256       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    274    274       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    288    288       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    292    292       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    309    309       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    330    330       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    351    351       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    360    360       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    375    375       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    432    432       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    466    466       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    480    480       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    508    508       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    529    529       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    553    553       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    560    560       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    580    580       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    612    612       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    650    650       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    665    665       N-linked (GlcNAc. . .) (Potential).
FT   LIPID       685    685       GPI-anchor amidated alanine (Potential).
FT   CONFLICT    320    320       Missing (in Ref. 4).
SQ   SEQUENCE   702 AA;  76795 MW;  6299AE26CDDBDB5C CRC64;

  Query Match             95.6%;  Score 43;  DB 1;  Length 702;
  Best Local Similarity   88.9%;  Pred. No. 6.6;
  Matches    8;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 YLSGANINL 9
              |||||:||
Db        605 YLSGANLNL 613


RESULT 3
Q8N4D0
ID   Q8N4D0      PRELIMINARY;      PRT;   702 AA.
AC   Q8N4D0;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   CEACAM5 protein.
OS   Homo sapiens (Human).
```

```
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Colon;
RX   MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA   Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
RA   Jones S.J., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length human
RT   and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Colon;
RA   Strausberg R.;
RL   Submitted (JUL-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC034671; AAH34671.1; -.
DR   HSSP; Q61353; 1L6Z.
DR   InterPro; IPR001589; Actbind_actnin.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR003598; Ig_c2.
DR   Pfam; PF00047; ig; 6.
DR   SMART; SM00408; IGc2; 3.
DR   PROSITE; PS00019; ACTININ_1; UNKNOWN_3.
DR   PROSITE; PS50835; IG_LIKE; 6.
SQ   SEQUENCE   702 AA;  76781 MW;  97CCFB7399A0B05A CRC64;

  Query Match             95.6%;  Score 43;  DB 2;  Length 702;
  Best Local Similarity   88.9%;  Pred. No. 6.6;
  Matches    8;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 YLSGANINL 9
              ||||||:||
Db        605 YLSGANLNL 613


RESULT 4
Q6AJS0
ID   Q6AJS0      PRELIMINARY;      PRT;   261 AA.
AC   Q6AJS0;
DT   25-OCT-2004 (TrEMBLrel. 28, Created)
DT   25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT   25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE   Probable flagellar hook basal body protein (FlgG).
GN   OrderedLocusNames=DP2681;
OS   Desulfotalea psychrophila.
OC   Bacteria; Proteobacteria; Deltaproteobacteria; Desulfobacterales;
OC   Desulfobulbaceae; Desulfotalea.
OX   NCBI_TaxID=84980;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=LSv54 / DSM 12343;
RX   PubMed=15305914;
RA   Rabus R., Ruepp A., Frickey T., Rattei T., Fartmann B., Stark M.,
RA   Bauer M., Zibat A., Lombardot T., Becker I., Amann J., Gellner K.,
RA   Teeling H., Leuschner W.D., Gloeckner F.-O., Lupas A.N., Amann R.,
RA   Klenk H.-P.;
RT   "The genome of Desulfotalea psychrophila, a sulfate-reducing bacterium
RT   from permanently cold Arctic sediments.";
RL   Environ. Microbiol. 6:887-902(2004).
DR   EMBL; CR522870; CAG37410.1; -.
DR   GO; GO:0009288; C:flagellum (sensu Bacteria); IEA.
DR   GO; GO:0003774; F:motor activity; IEA.
DR   GO; GO:0005198; F:structural molecule activity; IEA.
DR   GO; GO:0001539; P:ciliary or flagellar motility; IEA.
DR   InterPro; IPR010930; DUF1078.
DR   InterPro; IPR001444; Flag_bb_rod.
DR   Pfam; PF06429; DUF1078; 1.
DR   Pfam; PF00460; Flg_bb_rod; 1.
DR   PROSITE; PS00588; FLAGELLA_BB_ROD; 1.
KW   Complete proteome; Flagellum.
SQ   SEQUENCE   261 AA;  27650 MW;  36412CFBDC0DDBFC CRC64;

  Query Match             82.2%;  Score 37;  DB 2;  Length 261;
  Best Local Similarity   66.7%;  Pred. No. 40;
  Matches    6;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 YLSGANINL 9
              || |:|:||
Db        218 YLEGSNVNL 226


RESULT 5
Q73XB6
ID   Q73XB6      PRELIMINARY;      PRT;   265 AA.
AC   Q73XB6;
DT   05-JUL-2004 (TrEMBLrel. 27, Created)
DT   05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT   05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE   Hypothetical protein.
GN   OrderedLocusNames=MAP2393c;
OS   Mycobacterium paratuberculosis.
OC   Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC   Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1770;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=k10;
RA   Li L., Bannantine J., Zhang Q., Amonsin A., Alt D., Kapur V.;
RL   Submitted (SEP-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AE017235; AAS04710.1; -.
DR   GO; GO:0016491; F:oxidoreductase activity; IEA.
DR   GO; GO:0008152; P:metabolism; IEA.
DR   InterPro; IPR002198; ADH_short.
DR   InterPro; IPR002347; Adh_short_C2.
DR   Pfam; PF00106; adh_short; 1.
DR   PRINTS; PR00081; GDHRDH.
KW   Complete proteome.
SQ   SEQUENCE   265 AA;  27606 MW;  9210B08F3D422AC2 CRC64;

  Query Match             82.2%;  Score 37;  DB 2;  Length 265;
  Best Local Similarity   66.7%;  Pred. No. 41;
  Matches    6;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 YLSGANINL 9
              |::|||||:
Db        250 YMTGANINV 258


RESULT 6
Q9PKP0
ID   Q9PKP0      PRELIMINARY;      PRT;   506 AA.
AC   Q9PKP0;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Monooxygenase-related protein.
GN   OrderedLocusNames=TC0425;
OS   Chlamydia muridarum.
OC   Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydia.
```

```
OX   NCBI_TaxID=83560;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MoPn / Nigg;
RX   MEDLINE=20150255; PubMed=10684935; DOI=10.1093/nar/28.6.1397;
RA   Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
RA   White O., Hickey E.K., Peterson J.D., Utterback T.R., Berry K.J.,
RA   Bass S., Linher K.D., Weidman J.F., Khouri H.M., Craven B., Bowman C.,
RA   Dodson R.J., Gwinn M.L., Nelson W.C., DeBoy R.T., Kolonay J.F.,
RA   McClarty G., Salzberg S.L., Eisen J.A., Fraser C.M.;
RT   "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia
RT   pneumoniae AR39.";
RL   Nucleic Acids Res. 28:1397-1406(2000).
DR   EMBL; AE002309; AAF39281.1; -.
DR   PIR; C81704; C81704.
DR   TIGR; TC0425; -.
DR   GO; GO:0004497; F:monooxygenase activity; IEA.
DR   GO; GO:0006118; P:electron transport; IEA.
DR   InterPro; IPR003042; Rng_mnoxygenase.
DR   PRINTS; PR00420; RNGMNOXGNASE.
KW   Complete proteome; Monooxygenase.
SQ   SEQUENCE   506 AA;  57992 MW;  A5776CC63CB0C424 CRC64;

  Query Match             82.2%;  Score 37;  DB 2;  Length 506;
  Best Local Similarity   66.7%;  Pred. No. 79;
  Matches    6;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 YLSGANINL 9
              |||| |:|:
Db        290 YLSGVNLNI 298


RESULT 7
Q35903
ID   Q35903      PRELIMINARY;      PRT;   131 AA.
AC   Q35903;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   TRNA-Ser (Fragment).
OS   Strongylocentrotus pallidus.
OG   Mitochondrion.
OC   Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
OC   Echinoidea; Euechinoidea; Echinacea; Echinoida; Strongylocentrotidae;
OC   Strongylocentrotus.
OX   NCBI_TaxID=7670;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91175684; PubMed=2488275;
RA   Thomas W.K., Maa J., Wilson A.C.;
RT   "Shifting constraints on tRNA genes during mitochondrial DNA evolution
RT   in animals.";
RL   New Biol. 1:93-100(1989).
DR   EMBL; M27524; AAA32086.2; -.
DR   GO; GO:0005739; C:mitochondrion; IEA.
DR   GO; GO:0008137; F:NADH dehydrogenase (ubiquinone) activity; IEA.
DR   GO; GO:0042773; P:ATP synthesis coupled electron transport; IEA.
DR   InterPro; IPR001516; Oxidored_q1_N.
DR   Pfam; PF00662; Oxidored_q1_N; 1.
KW   Mitochondrion.
FT   NON_TER     131    131
SQ   SEQUENCE   131 AA;  14405 MW;  903318CDAD6E4C0D CRC64;

  Query Match             80.0%;  Score 36;  DB 2;  Length 131;
  Best Local Similarity   77.8%;  Pred. No. 32;
  Matches    7;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 YLSGANINL 9
              ||| :||||
Db         31 YLSNSNINL 39


RESULT 8
Q85I11
ID   Q85I11      PRELIMINARY;      PRT;   253 AA.
AC   Q85I11;
DT   01-JUN-2003 (TrEMBLrel. 24, Created)
DT   01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   ATPase 6 (Fragment).
GN   Name=ATP6;
OS   Strongylocentrotus pallidus.
OG   Mitochondrion.
OC   Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
OC   Echinoidea; Euechinoidea; Echinacea; Echinoida; Strongylocentrotidae;
OC   Strongylocentrotus.
OX   NCBI_TaxID=7670;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=22707965; PubMed=12823452;
RA   Biermann C.H., Kessing B.D., Palumbi S.R.;
RT   "Phylogeny and development of marine model species: strongylocentrotid
RT   sea urchins.";
RL   Evol. Dev. 5:360-371(2003).
CC   -!- FUNCTION: Key component of the proton channel; it may play a
CC       direct role in the translocation of protons across the membrane
CC       (By similarity).
CC   -!- CATALYTIC ACTIVITY: ATP + H(2)O + H(+)(In) = ADP + phosphate +
CC       H(+)(Out).
CC   -!- SUBUNIT: F-type ATPases have 2 components, CF(1) - the catalytic
CC       core - and CF(0) - the membrane proton channel. CF(1) has five
CC       subunits: alpha(3), beta(3), gamma(1), delta(1), epsilon(1). CF(0)
CC       has three main subunits: a, b and c (By similarity).
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein (By similarity).
CC   -!- SIMILARITY: Belongs to the ATPase A chain family.
DR   EMBL; AY221013; AAP21723.1; -.
DR   GO; GO:0005739; C:mitochondrion; IEA.
DR   GO; GO:0016469; C:proton-transporting two-sector ATPase complex; IEA.
DR   GO; GO:0016820; F:hydrolase activity, acting on acid anhydrid. . .; IEA.
DR   GO; GO:0008137; F:NADH dehydrogenase (ubiquinone) activity; IEA.
DR   GO; GO:0042773; P:ATP synthesis coupled electron transport; IEA.
DR   GO; GO:0006811; P:ion transport; IEA.
DR   GO; GO:0015992; P:proton transport; IEA.
DR   InterPro; IPR000568; ATPsynt_Asub.
DR   InterPro; IPR001516; Oxidored_q1_N.
DR   Pfam; PF00119; ATP-synt_A; 1.
DR   Pfam; PF00662; Oxidored_q1_N; 1.
DR   TIGRFAMs; TIGR01131; ATP_synt_6_or_A; 1.
KW   CF(0); Hydrogen ion transport; Ion transport; Mitochondrion;
KW   Transmembrane; Transport.
FT   NON_TER     253    253
SQ   SEQUENCE   253 AA;  28159 MW;  D55183040672A77D CRC64;

  Query Match             80.0%;  Score 36;  DB 2;  Length 253;
  Best Local Similarity   77.8%;  Pred. No. 62;
  Matches    7;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 YLSGANINL 9
              ||| :||||
Db        153 YLSNSNINL 161


RESULT 9
Q81CR9
ID   Q81CR9      PRELIMINARY;      PRT;   281 AA.
AC   Q81CR9;
DT   01-JUN-2003 (TrEMBLrel. 24, Created)
DT   01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   L-alanyl-D-glutamate peptidase (EC 3.4.-.-).
GN   OrderedLocusNames=BC2677;
OS   Bacillus cereus (strain ATCC 14579 / DSM 31).
OC   Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX   NCBI_TaxID=226900;
RN   [1]
RP   SEQUENCE FROM N.A.
```

```
RX   MEDLINE=22608415; PubMed=12721630; DOI=10.1038/nature01582;
RA   Ivanova N., Sorokin A., Anderson I., Galleron N., Candelon B.,
RA   Kapatral V., Bhattacharyya A., Reznik G., Mikhailova N., Lapidus A.,
RA   Chu L., Mazur M., Goltsman E., Larsen N., D'Souza M., Walunas T.,
RA   Grechkin Y., Pusch G., Haselkorn R., Fonstein M., Ehrlich S.D.,
RA   Overbeek R., Kyrpides N.C.;
RT   "Genome sequence of Bacillus cereus and comparative analysis with
RT   Bacillus anthracis.";
RL   Nature 423:87-91(2003).
DR   EMBL; AE017006; AAP09633.1; -.
DR   GO; GO:0016787; F:hydrolase activity; IEA.
DR   InterPro; IPR009045; Hedgehog_sig_N.
DR   InterPro; IPR003646; SH3_bac.
DR   SMART; SM00287; SH3b; 2.
KW   Complete proteome; Hydrolase.
SQ   SEQUENCE   281 AA;  31402 MW;  9E42473A2ACCD71E CRC64;

  Query Match             80.0%;  Score 36;  DB 2;  Length 281;
  Best Local Similarity   55.6%;  Pred. No. 69;
  Matches    5;  Conservative    4;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 YLSGANINL 9
              |::|:|:||
Db        151 YINGSNVNL 159


RESULT 10
Q7DAC6
ID   Q7DAC6      PRELIMINARY;      PRT;   518 AA.
AC   Q7DAC6;
DT   05-JUL-2004 (TrEMBLrel. 27, Created)
DT   05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT   05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE   Substrate--CoA ligase.
GN   OrderedLocusNames=MT0175;
OS   Mycobacterium tuberculosis.
OC   Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC   Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1773;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CDC 1551 / Oshkosh;
RX   MEDLINE=22206494; PubMed=12218036;
RX   DOI=10.1128/JB.184.19.5479-5490.2002;
RA   Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA   Peterson J.D., DeBoy R.T., Dodson R.J., Gwinn M.L., Haft D.H.,
RA   Hickey E.K., Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D.,
RA   Salzberg S.L., Delcher A., Utterback T.R., Weidman J.F., Khouri H.M.,
RA   Gill J., Mikula A., Bishai W., Jacobs W.R. Jr., Venter J.C.,
RA   Fraser C.M.;
RT   "Whole-genome comparison of Mycobacterium tuberculosis clinical and
RT   laboratory strains.";
RL   J. Bacteriol. 184:5479-5490(2002).
CC   -!- SIMILARITY: Belongs to the ATP-dependent AMP-binding enzyme
CC       family.
DR   EMBL; AE000516; AAK44395.1; -.
DR   HSSP; P08659; 1BA3.
DR   TIGR; MT0175; -.
DR   GO; GO:0016874; F:ligase activity; IEA.
DR   GO; GO:0008152; P:metabolism; IEA.
DR   InterPro; IPR000873; AMP-bind.
DR   Pfam; PF00501; AMP-binding; 1.
DR   PRINTS; PR00154; AMPBINDING.
DR   PROSITE; PS00455; AMP_BINDING; 1.
KW   Ligase.
SQ   SEQUENCE   518 AA;  55708 MW;  AC08041D2F8C9C3A CRC64;

  Query Match             80.0%;  Score 36;  DB 2;  Length 518;
  Best Local Similarity   87.5%;  Pred. No. 1.3e+02;
  Matches    7;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 YLSGANIN 8
              | ||||||
Db        181 YTSGANIN 188


RESULT 11
O07411
ID   O07411      PRELIMINARY;      PRT;   554 AA.
AC   O07411;
DT   01-JUL-1997 (TrEMBLrel. 04, Created)
DT   01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   PROBABLE FATTY-ACID-CoA LIGASE FADD5 (FATTY-ACID-CoA SYNTHETASE)
DE   (FATTY-ACID-CoA SYNTHASE) (EC 6.2.1.-).
GN   Name=fadD5; OrderedLocusNames=Rv0166;
OS   Mycobacterium tuberculosis.
OC   Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC   Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1773;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37Rv;
RX   MEDLINE=98295987; PubMed=9634230; DOI=10.1038/31159;
RA   Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C.M.,
RA   Harris D.E., Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III,
RA   Tekaia F., Badcock K., Basham D., Brown D., Chillingworth T.,
RA   Connor R., Davies R.M., Devlin K., Feltwell T., Gentles S., Hamlin N.,
RA   Holroyd S., Hornsby T., Jagels K., Krogh A., McLean J., Moule S.,
RA   Murphy L.D., Oliver S., Osborne J., Quail M.A., Rajandream M.A.,
RA   Rogers J., Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA   Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT   "Deciphering the biology of Mycobacterium tuberculosis from the
RT   complete genome sequence.";
RL   Nature 393:537-544(1998).
CC   -!- SIMILARITY: Belongs to the ATP-dependent AMP-binding enzyme
CC       family.
DR   EMBL; BX842572; CAB09749.1; -.
DR   PIR; A70904; A70904.
DR   HSSP; P08659; 1LCI.
DR   TubercuList; Rv0166; -.
DR   GO; GO:0016874; F:ligase activity; IEA.
DR   GO; GO:0008152; P:metabolism; IEA.
DR   InterPro; IPR000873; AMP-bind.
DR   Pfam; PF00501; AMP-binding; 1.
DR   PRINTS; PR00154; AMPBINDING.
DR   PROSITE; PS00455; AMP_BINDING; 1.
KW   Complete proteome; Ligase.
SQ   SEQUENCE   554 AA;  59905 MW;  3AC3100FAF0B9E88 CRC64;

  Query Match             80.0%;  Score 36;  DB 2;  Length 554;
  Best Local Similarity   87.5%;  Pred. No. 1.4e+02;
  Matches    7;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 YLSGANIN 8
              | ||||||
Db        217 YTSGANIN 224


RESULT 12
Q7U2P2
ID   Q7U2P2      PRELIMINARY;      PRT;   554 AA.
AC   Q7U2P2;
DT   01-OCT-2003 (TrEMBLrel. 25, Created)
DT   01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   PROBABLE FATTY-ACID-CoA LIGASE FADD5 (FATTY-ACID-CoA SYNTHETASE)
DE   (FATTY-ACID-CoA SYNTHASE) (EC 6.2.1.-).
GN   Name=fadD5; OrderedLocusNames=Mb0172;
OS   Mycobacterium bovis.
OC   Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC   Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1765;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=AF2122/97;
```

```
RX   MEDLINE=22709107; PubMed=12788972; DOI=10.1073/pnas.1130426100;
RA   Garnier T., Eiglmeier K., Camus J.-C., Medina N., Mansoor H.,
RA   Pryor M., Duthoy S., Grondin S., Lacroix C., Monsempe C., Simon S.,
RA   Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,
RA   Parkhill J., Barrell B.G., Cole S.T., Gordon S.V., Hewinson R.G.;
RT   "The complete genome sequence of Mycobacterium bovis.";
RL   Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882(2003).
CC   -!- SIMILARITY: Belongs to the ATP-dependent AMP-binding enzyme
CC       family.
DR   EMBL; BX248334; CAD93036.1; -.
DR   HSSP; P08659; 1BA3.
DR   GO; GO:0016874; F:ligase activity; IEA.
DR   GO; GO:0008152; P:metabolism; IEA.
DR   InterPro; IPR000873; AMP-bind.
DR   Pfam; PF00501; AMP-binding; 1.
DR   PROSITE; PS00455; AMP_BINDING; UNKNOWN_1.
KW   Complete proteome; Ligase.
SQ   SEQUENCE   554 AA;  59861 MW;  D5E984AFA1AB7AF6 CRC64;

  Query Match             80.0%;  Score 36;  DB 2;  Length 554;
  Best Local Similarity   87.5%;  Pred. No. 1.4e+02;
  Matches    7;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 YLSGANIN 8
              | ||||||
Db        217 YTSGANIN 224


RESULT 13
Q6BHN8
ID   Q6BHN8      PRELIMINARY;      PRT;   824 AA.
AC   Q6BHN8;
DT   25-OCT-2004 (TrEMBLrel. 28, Created)
DT   25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT   25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE   Similar to CA0225|CaNUP84 Candida albicans CaNUP84.
GN   ORFNames=DEHA0G18227g;
OS   Debaryomyces hansenii CBS767.
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Debaryomyces.
OX   NCBI_TaxID=284592;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CBS767;
RG   Genolevures;
RA   Dujon B., Sherman D., Fischer G., Durrens P., Casaregola S.,
RA   Lafontaine I., de Montigny J., Marck C., Neuveglise C., Talla E.,
RA   Goffard N., Frangeul L., Aigle M., Anthouard V., Babour A., Barbe V.,
RA   Barnay S., Blanchin S., Beckerich J.M., Beyne E., Bleykasten C.,
RA   Boisrame A., Boyer J., Cattolico L., Confanioleri F., de Daruvar A.,
RA   Despons L., Fabre E., Fairhead C., Ferry-Dumazet H., Groppi A.,
RA   Hantraye F., Hennequin C., Jauniaux N., Joyet P., Kachouri R.,
RA   Kerrest A., Koszul R., Lemaire M., Lesur I., Ma L., Muller H.,
RA   Nicaud J.M., Nikolski M., Oztas S., Ozier-Kalogeropoulos O.,
RA   Pellenz S., Potier S., Richard G.F., Straub M.L., Suleau A.,
RA   Swennene D., Tekaia F., Wesolowski-Louvel M., Westhof E., Wirth B.,
RA   Zeniou-Meyer M., Zivanovic I., Bolotin-Fukuhara M., Thierry A.,
RA   Bouchier C., Caudron B., Scarpelli C., Gaillardin C., Weissenbach J.,
RA   Wincker P., Souciet J.L.;
RT   "Genome evolution in yeasts.";
RL   Nature 430:35-44(2004).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CBS767;
RA   Genoscope;
RL   Submitted (JUL-2004) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; CR382139; CAG90785.1; -.
DR   GO; GO:0005643; C:nuclear pore; IEA.
DR   GO; GO:0006810; P:transport; IEA.
DR   InterPro; IPR007252; Nup84_Nup100.
DR   Pfam; PF04121; Nup84_Nup100; 1.
SQ   SEQUENCE   824 AA;  94352 MW;  1D8D892280D0068F CRC64;

  Query Match             80.0%;  Score 36;  DB 2;  Length 824;
  Best Local Similarity   77.8%;  Pred. No. 2.1e+02;
  Matches    7;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 YLSGANINL 9
              |||| ||:|
Db        312 YLSGGNISL 320


RESULT 14
Q7RTZ4
ID   Q7RTZ4      PRELIMINARY;      PRT;   159 AA.
AC   Q7RTZ4;
DT   01-MAR-2004 (TrEMBLrel. 26, Created)
DT   01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Intracellular interleukin-1 receptor antagonist (IcIL-1ra).
GN   Name=IL1RN (IL1F3);
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20545212; PubMed=11093146;
RX   DOI=10.1002/1521-4141(200011)30:11<3299::AID-IMMU3299>3.0.CO;2-S;
RA   Barton J.L., Herbst R., Bosisio D., Higgins L., Nicklin M.J.;
RT   "A tissue specific IL-1 receptor antagonist homolog from the IL-1
RT   cluster lacks IL-1, IL-1ra, IL-18 and IL-18 antagonist activities.";
RL   Eur. J. Immunol. 30:3299-3308(2000).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97312693; PubMed=9169134; DOI=10.1006/geno.1997.4654;
RA   Nothwang H.G., Strahm B., Denich D., Kuebler M., Schwabe J.,
RA   Gingrich J.C., Jauch A., Cox A., Nicklin M.J.H., Kurnit D.M.,
RA   Hildebrandt F.;
RT   "Molecular cloning of the interleukin-1 gene cluster: construction of
RT   an integrated YAC/PAC contig and a partial transcriptional map in the
RT   region of chromosome 2q13.";
RL   Genomics 41:370-378(1997).
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94245215; PubMed=8188271;
RA   Nicklin M.J.H., Weith A., Duff G.W.;
RT   "A Physical map of the region encompassing the human interleukin-1-
RT   alpha, interleukin-1-beta and interleukin-1 receptor genes.";
RL   Genomics 19:382-384(1994).
RN   [4]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21988050; PubMed=11991722; DOI=10.1006/geno.2002.6751;
RA   Nicklin M.J.H., Barton J.L., Nguyen M., Fitzgerald M.G., Duff W.G.,
RA   Kornman K.;
RT   "A sequence-based map of the nine genes of the human interleukin-1
RT   cluster.";
RL   Genomics 79:718-725(2002).
RN   [5]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99443727; PubMed=10512743; DOI=10.1006/bbrc.1999.1440;
RA   Mulero J.J., Pace A.M., Nelken S.T., Loeb D.D., Correa T.R.,
RA   Drmanac R., Ford J.E.;
RT   "IL1HY1: A novel interleukin-1 receptor antagonist gene.";
RL   Biochem. Biophys. Res. Commun. 263:702-706(1999).
RN   [6]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20092888; PubMed=10625660; DOI=10.1074/jbc.275.2.1169;
RA   Smith D.E., Renshaw B.R., Ketchem R.R., Kubin M., Garka K.E.,
RA   Sims J.E.;
RT   "Four new members expand the interleukin-1 superfamily.";
RL   J. Biol. Chem. 275:1169-1175(2000).
RN   [7]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20209405; PubMed=10744718; DOI=10.1074/jbc.275.14.10308;
RA   Kumar S., McDonnell P.C., Lehr R., Tierney L., Tzimas M.N.,
```

```
RA   Griswold D.E., Capper E.A., Tal-Singer R., Wells G.I., Doyle M.L.,
RA   Young P.R.;
RT   "Identification and initial characterization of four novel members of
RT   the interleukin-1 family.";
RL   J. Biol. Chem. 275:10308-10314(2000).
RN   [8]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20318623; PubMed=10860666; DOI=10.1006/geno.2000.6184;
RA   Busfield S.J., Comrack C.A., Yu G., Chickering T.W., Smutko J.S.,
RA   Zhou H., Leiby K.R., Holmgren L.M., Gearing D.P., Pan Y.;
RT   "Identification and gene organization of three novel members of the
RT   IL-1 family on human chromosome 2.";
RL   Genomics 66:213-216(2000).
RN   [9]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21066552; PubMed=11145836; DOI=10.1006/cyto.2000.0799;
RA   Pan G., Risser P., Mao W., Baldwin D.T., Zhong A.W., Filvaroff E.,
RA   Yansura D., Lewis L., Eigenbrot C., Henzel W.J., Vandlen R.;
RT   "IL-1H, an interleukin 1-related protein that binds IL-18 receptor/IL-
RT   1Rrp.";
RL   Cytokine 13:1-7(2001).
RN   [10]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21282953; PubMed=11278614; DOI=10.1074/jbc.M010095200;
RA   Lin H.S., Ho A.S., Haley-Vicente D., Zhang J., Bernal-Fussel J.,
RA   Pace A.M., Hansen D., Schweighofer K., Mize N.K., Ford J.E.;
RT   "Cloning and characterization of IL-1HY2, a novel interleukin-1 family
RT   member.";
RL   J. Biol. Chem. 276:20597-20602(2001).
RN   [11]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21359532; PubMed=11466363;
RA   Debets R., Timans J.C., Homey B., Zurawski S., Sana T.R., Lo S.,
RA   Wagner J., Edwards G., Clifford T., Menon S., Bazan J.F.,
RA   Kastelein R.A.;
RT   "Two novel IL-1 family members, IL-1 delta and IL-1 epsilon, function
RT   as an antagonist and agonist of NF-kB activation through the orphan
RT   IL-1 receptor-related protein 2.J. Immunol. 167: 1440-1446.";
RL   J. Immunol. 167:1440-1446(2001).
RN   [12]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21459116; PubMed=11574262; DOI=10.1016/S1471-4906(01)02040-3;
RA   Sims J.E., Nicklin M.J., Bazan J.F., Barton J.L., Busfield S.J.,
RA   Ford J.E., Kastelein R.A., Kumar S., Lin H., Mulero J.J., Pan G.,
RA   Pan Y., Smith D.E., Young P.R.;
RT   "A new nomenclature for the IL-1-family genes.";
RL   Trends Immunol. 22:536-537(2001).
CC   -!- MISCELLANEOUS: The sequence shown here is derived from an
CC       EMBL/GenBank/DDBJ third party annotation (TPA) entry.
CC   -!- SIMILARITY: Belongs to the IL-1 family.
DR   EMBL; BN000002; CAD29879.1; -.
DR   HSSP; P18510; 1IRP.
DR   GO; GO:0005576; C:extracellular; IEA.
DR   GO; GO:0005152; F:interleukin-1 receptor antagonist activity; IEA.
DR   GO; GO:0004872; F:receptor activity; IEA.
DR   GO; GO:0006955; P:immune response; IEA.
DR   InterPro; IPR008996; Cytok_IL1_like.
DR   InterPro; IPR003297; InterleukinIL1RA.
DR   InterPro; IPR000975; Interleukin_1.
DR   Pfam; PF00340; IL1; 1.
DR   PRINTS; PR00264; INTERLEUKIN1.
DR   PRINTS; PR01360; INTRLEUKIN1X.
DR   ProDom; PD002536; Interleukin_1; 1.
DR   PROSITE; PS00253; INTERLEUKIN_1; 1.
KW   Receptor.
SQ   SEQUENCE   159 AA;  17888 MW;  C1D66CDF0D2F7B44 CRC64;

  Query Match             77.8%;  Score 35;  DB 2;  Length 159;
  Best Local Similarity   66.7%;  Pred. No. 62;
  Matches    6;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 YLSGANINL 9
              || | |:||
Db         41 YLQGPNVNL 49


RESULT 15
IL1X_HUMAN
ID   IL1X_HUMAN     STANDARD;      PRT;   177 AA.
AC   P18510; Q14628; Q96GD6; Q9UPC0;
DT   01-NOV-1990 (Rel. 16, Created)
DT   01-NOV-1990 (Rel. 16, Last sequence update)
DT   25-OCT-2004 (Rel. 45, Last annotation update)
DE   Interleukin-1 receptor antagonist protein precursor (IL-1ra) (IRAP)
DE   (IL1 inhibitor) (IL-1RN) (ICIL-1RA).
GN   Name=IL1RN; Synonyms=IL1RA;
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORM 1).
RX   MEDLINE=90220867; PubMed=2139180; DOI=10.1038/344633a0;
RA   Carter D.B., Deibel M.R. Jr., Dunn C.J., Tomich C.S.C., Laborde A.L.,
RA   Slightom J.L., Berger A.E., Bienkowski M.J., Sun F.F., McEwan R.N.,
RA   Harris P.K.W., Yem A.W., Waszak G.A., Chosay J.G., Sieu L.C.,
RA   Hardee M.M., Zurcher-Neely H.A., Reardon I.M., Heinrikson R.L.,
RA   Truesdell S.E., Shelly J.A., Eessalu T.E., Taylor B.M., Tracey D.E.;
RT   "Purification, cloning, expression and biological characterization of
RT   an interleukin-1 receptor antagonist protein.";
RL   Nature 344:633-638(1990).
RN   [2]
RP   SEQUENCE FROM N.A. (ISOFORM 1).
RX   MEDLINE=90136921; PubMed=2137201; DOI=10.1038/343341a0;
RA   Eisenberg S.P., Evans R.J., Arend W.P., Verderber E., Brewer M.T.,
RA   Hannum C.H., Thompson R.C.;
RT   "Primary structure and functional expression from complementary DNA of
RT   a human interleukin-1 receptor antagonist.";
RL   Nature 343:341-346(1990).
RN   [3]
RP   SEQUENCE FROM N.A. (ISOFORM 1).
RX   MEDLINE=91271363; PubMed=1828896;
RA   Eisenberg S.P., Brewer M.T., Verderber E., Heimdal P.,
RA   Brandhuber B.J., Thompson R.C.;
RT   "Interleukin 1 receptor antagonist is a member of the interleukin 1
RT   gene family: evolution of a cytokine control mechanism.";
RL   Proc. Natl. Acad. Sci. U.S.A. 88:5232-5236(1991).
RN   [4]
RP   SEQUENCE FROM N.A. (ISOFORM 1).
RX   MEDLINE=92338323; PubMed=1385987; DOI=10.1016/1043-4666(92)90041-O;
RA   Lennard A., Gorman P., Carrier M., Griffiths S., Scotney H., Sheer D.,
RA   Solari R.;
RT   "Cloning and chromosome mapping of the human interleukin-1 receptor
RT   antagonist gene.";
RL   Cytokine 4:83-89(1992).
RN   [5]
RP   SEQUENCE FROM N.A. (ISOFORMS 1 AND 3).
RX   MEDLINE=97146044; PubMed=8992991;
RA   Jenkins J.K., Drong R.F., Shuck M.E., Bienkowski M.J., Slightom J.L.,
RA   Arend W.P., Smith M.F. Jr.;
RT   "Intracellular IL-1 receptor antagonist promoter: cell type-specific
RT   and inducible regulatory regions.";
RL   J. Immunol. 158:748-755(1997).
RN   [6]
RP   SEQUENCE FROM N.A. (ISOFORM 2).
RX   MEDLINE=91219436; PubMed=1827201;
RA   Haskill S., Martin G., van Le L., Morris J., Peace A., Bigler C.F.,
RA   Jaffe G.J., Hammerberg C., Sporn S.A., Fong S., Arend W.P., Ralph P.;
RT   "cDNA cloning of an intracellular form of the human interleukin 1
RT   receptor antagonist associated with epithelium.";
RL   Proc. Natl. Acad. Sci. U.S.A. 88:3681-3685(1991).
RN   [7]
RP   SEQUENCE FROM N.A. (ISOFORM 3).
RX   MEDLINE=95355865; PubMed=7629520;
RA   Muzio M., Polentarutti N., Sironi M., Poli G., De Gioia L.,
RA   Introna M., Mantovani A., Colotta F.;
```

```
RT   "Cloning and characterization of a new isoform of the interleukin 1
RT   receptor antagonist.";
RL   J. Exp. Med. 182:623-628(1995).
RN   [8]
RP   SEQUENCE FROM N.A. (ISOFORM 2).
RA   Rieder M.J., Carrington D.P., Hastings N.C., Ahearn M.O.,
RA   Kuldanek S.A., Rajkumar N., Toth E.J., Yi Q., Nickerson D.A.;
RT   "SeattleSNPs. NHLBI HL66682 program for genomic applications, UW-
RT   FHCRC, Seattle, WA (URL: http://pga.gs.washington.edu).";
RL   Submitted (DEC-2002) to the EMBL/GenBank/DDBJ databases.
RN   [9]
RP   SEQUENCE FROM N.A. (ISOFORM 2).
RC   TISSUE=Pancreas;
RX   MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length human
RT   and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [10]
RP   SEQUENCE OF 26-45, AND CARBOHYDRATE-LINKAGE SITE ASN-109.
RX   MEDLINE=90136920; PubMed=2137200; DOI=10.1038/343336a0;
RA   Hannum C.H., Wilcox C.J., Arend W.P., Joslin F.G., Dripps D.J.,
RA   Heimdal P.L., Armes L.G., Sommer A., Eisenberg S.P., Thompson R.C.;
RT   "Interleukin-1 receptor antagonist activity of a human interleukin-1
RT   inhibitor.";
RL   Nature 343:336-340(1990).
RN   [11]
RP   SEQUENCE OF 26-52.
RX   MEDLINE=90354444; PubMed=2143761;
RA   Bienkowski M.J., Eessalu T.E., Berger A.E., Truesdell S.E.,
RA   Shelly J.A., Laborde A.L., Zurcher-Neely H.A., Reardon I.M.,
RA   Heinrikson R.L., Chosay J.G., Tracey D.E.;
RT   "Purification and characterization of interleukin 1 receptor level
RT   antagonist proteins from THP-1 cells.";
RL   J. Biol. Chem. 265:14505-14511(1990).
RN   [12]
RP   SEQUENCE OF 35-177 FROM N.A. (ISOFORM 4).
RX   MEDLINE=98183404; PubMed=9514884; DOI=10.1006/bbrc.1998.8217;
RA   Weissbach L., Tran K., Colquhoun S.A., Champliaud M.F., Towle C.A.;
RT   "Detection of an interleukin-1 intracellular receptor antagonist mRNA
RT   variant.";
RL   Biochem. Biophys. Res. Commun. 244:91-95(1998).
RN   [13]
RP   STRUCTURE BY NMR.
RX   MEDLINE=92297633; PubMed=1534997;
RA   Stockman B.J., Scahill T.A., Roy M., Ulrich E.L., Strakalaitis N.A.,
RA   Brunner D.P., Yem A.W., Deibel M.R. Jr.;
RT   "Secondary structure and topology of interleukin-1 receptor antagonist
RT   protein determined by heteronuclear three-dimensional NMR
RT   spectroscopy.";
RL   Biochemistry 31:5237-5244(1992).
RN   [14]
RP   STRUCTURE BY NMR.
RX   MEDLINE=94320651; PubMed=8045306; DOI=10.1016/0014-5793(94)00643-1;
RA   Stockman B.J., Scahill T.A., Strakalaitis N.A., Brunner D.P.,
RA   Yem A.W., Deibel M.R. Jr.;
RT   "Solution structure of human interleukin-1 receptor antagonist
RT   protein.";
RL   FEBS Lett. 349:79-83(1994).
RN   [15]
RP   X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).
RX   MEDLINE=94230368; PubMed=8175703;
RA   Vigers G.P.A., Caffes P., Evans R.J., Thompson R.C., Eisenberg S.P.,
RA   Brandhuber B.J.;
RT   "X-ray structure of interleukin-1 receptor antagonist at 2.0-A
RT   resolution.";
RL   J. Biol. Chem. 269:12874-12879(1994).
RN   [16]
RP   X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS).
RX   MEDLINE=95172072; PubMed=7867645;
RA   Schreuder H.A., Rondeau J.-M., Tardif C., Soffientini A., Sarubbi E.,
RA   Akeson A., Bowlin T.L., Yanofsky S., Barrett R.W.;
RT   "Refined crystal structure of the interleukin-1 receptor antagonist.
RT   Presence of a disulfide link and a cis-proline.";
RL   Eur. J. Biochem. 227:838-847(1995).
RN   [17]
RP   X-RAY CRYSTALLOGRAPHY (2.7 ANGSTROMS) OF 32-177 IN COMPLEX WITH IL1R.
RX   MEDLINE=97215904; PubMed=9062194;
RA   Schreuder H., Tardif C., Trump-Kallmeyer S., Soffientini A.,
RA   Sarubbi E., Akeson A., Bowlin T., Yanofsky S., Barrett R.W.;
RT   "A new cytokine-receptor binding mode revealed by the crystal
RT   structure of the IL-1 receptor with an antagonist.";
RL   Nature 386:194-200(1997).
CC   -!- FUNCTION: Inhibits the activity of IL-1 by binding to its
CC       receptor. Has no IL-1 like activity.
CC   -!- SUBCELLULAR LOCATION: Secreted (isoform 1). Cytoplasmic (isoforms
CC       2, 3 and 4).
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=4;
CC       Name=1;
CC         IsoId=P18510-1; Sequence=Displayed;
CC       Name=2; Synonyms=icIL-1ra;
CC         IsoId=P18510-2; Sequence=VSP_002649;
CC       Name=3; Synonyms=icIL-1ra type II;
CC         IsoId=P18510-3; Sequence=VSP_002650;
CC       Name=4;
CC         IsoId=P18510-4; Sequence=VSP_002651;
CC   -!- TISSUE SPECIFICITY: The intracellular form of IL1RN is
CC       predominantly expressed in epithelial cells.
CC   -!- SIMILARITY: Belongs to the IL-1 family.
CC   -!- DATABASE: NAME=R&D Systems' cytokine source book: IL1RN;
CC       WWW="http://www.rndsystems.com/asp/g_sitebuilder.asp?bodyId=205".
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M55646; AAA59138.1; -.
DR   EMBL; M63099; AAB41943.1; -.
DR   EMBL; X52015; CAA36262.1; -.
DR   EMBL; X53296; CAA37386.1; -.
DR   EMBL; X64532; CAA45832.1; -.

  Query Match             77.8%;  Score 35;  DB 1;  Length 177;
  Best Local Similarity   66.7%;  Pred. No. 69;
  Matches    6;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 YLSGANINL 9
              || | |:||
Db         59 YLQGPNVNL 67


Search completed: May 17, 2005, 06:23:38
Job time : 53.75 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: May 17, 2005, 06:13:14 ; Search time 66 Seconds
(without alignments)
52.740 Million cell updates/sec

Title: US-10-725-373-4
Perfect score: 45
Sequence: 1 YLSGANINL 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_16Dec04:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	45	100.0	9	2	AAY09528	Aay09528 Carcinoem
2	43	95.6	9	2	AAW39723	Aaw39723 Human car
3	43	95.6	9	2	AAW70045	Aaw70045 CEA deriv
4	43	95.6	9	2	AAW77134	Aaw77134 CEA synth
5	43	95.6	9	2	AAY47655	Aay47655 Immunogen
6	43	95.6	9	2	AAY09525	Aay09525 Carcinoem
7	43	95.6	9	3	AAB13749	Aab13749 Peptide f
8	43	95.6	9	4	AAE02673	Aae02673 Human CEA
9	43	95.6	9	4	AAE00463	Aae00463 Human tum
10	43	95.6	9	4	AAE05123	Aae05123 Carcinoem
11	43	95.6	9	4	AAB82776	Aab82776 Carcinoem
12	43	95.6	9	5	ABG79073	Abg79073 Human CEA
13	43	95.6	9	5	AAE26805	Aae26805 Human HLA
14	43	95.6	9	5	AAU95893	Aau95893 Immunogen
15	43	95.6	9	5	AAE19088	Aae19088 HLA-A24 r
16	43	95.6	9	6	ABR56428	Abr56428 CEA epito
17	43	95.6	9	6	ABP98779	Abp98779 CAE pepti
18	43	95.6	9	6	ABR44529	Abr44529 CEA epito
19	43	95.6	9	7	ADD84715	Add84715 Human car
20	43	95.6	9	7	AAO24210	Aao24210 Human tum
21	43	95.6	9	8	ADG20333	Adg20333 Antigenic
22	43	95.6	9	8	ADJ36382	Adj36382 CEA epito
23	43	95.6	9	8	ADM12344	Adm12344 MHC class
24	43	95.6	9	8	ADM12341	Adm12341 MHC class
25	43	95.6	9	8	ADM72999	Adm72999 Human CEA
26	43	95.6	9	8	ADL46188	Adl46188 Human CAP
27	43	95.6	9	8	ADO38561	Ado38561 Carcinoem
28	43	95.6	9	8	ADO38564	Ado38564 Carcinoem
29	43	95.6	10	2	AAY46555	Aay46555 Immunogen
30	43	95.6	10	5	AAU11587	Aau11587 Human car
31	43	95.6	10	6	ABR83489	Abr83489 Human car
32	43	95.6	10	8	ADM72998	Adm72998 Human CEA
33	43	95.6	10	8	ADP80031	Adp80031 Human HLA
34	43	95.6	14	4	AAB88124	Aab88124 CD66 pept
35	43	95.6	25	5	AAU82083	Aau82083 T-cell sp
36	43	95.6	27	5	AAU82075	Aau82075 T-cell sp
37	43	95.6	84	7	ADE76434	Ade76434 Human CEA
38	43	95.6	102	8	ADP80496	Adp80496 Human epi
39	43	95.6	107	2	AAW86133	Aaw86133 Protein s
40	43	95.6	107	7	AAO24248	Aao24248 708 anti-
41	43	95.6	178	1	AAP93499	Aap93499 Sequence
42	43	95.6	227	5	AAE24334	Aae24334 Human lun
43	43	95.6	372	6	ABU04802	Abu04802 Human exp
44	43	95.6	468	2	AAR77436	Aar77436 BGP (1-31
45	43	95.6	493	2	AAR77435	Aar77435 BGP (1-31

ALIGNMENTS

RESULT 1
AAY09528
ID AAY09528 standard; peptide; 9 AA.
XX
AC AAY09528;
XX
DT 20-JUL-1999 (first entry)
XX
DE Carcinoembryonic antigen peptide agonist SEQ ID NO:4.
XX
KW Carcinoembryonic antigen; CEA; human; agonist; antagonist;
KW immune response; carcinoma; gastrointestinal; breast; pancreatic;
KW bladder; ovarian; lung; prostatic; T cell proliferation; cancer;
KW adoptive transfer therapy; autoimmune reaction; immunotherapy.
XX
OS Homo sapiens.
OS Synthetic.
XX
PN WO9919478-A1.
XX
PD 22-APR-1999.
XX
PF 22-SEP-1998; 98WO-US019794.
XX
PR 10-OCT-1997; 97US-0061589P.
XX
PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
XX
PI Schlom J, Barzaga E, Zaremba S;
XX
DR WPI; 1999-326544/27.
XX
PT Peptide agonists and antagonists of carcinoembryonal antigen.
XX
PS Claim 5; Page 53; 72pp; English.
XX
CC The present invention describes peptides (A) that comprise agonists (Ia)
CC or antagonists (Ib) of human carcinoembryonal antigen (CEA). (Ia) are
CC used in vaccines to kill or inhibit carcinoma cells that express CEA or
CC its epitopes, particularly for treating gastrointestinal, breast,
CC pancreatic, bladder, ovarian, lung or prostatic carcinoma. They can also
CC be used to proliferate T cells, e.g. from vaccinated subjects, for use in
CC adoptive transfer therapy. (Ib) are used to inhibit CEA-specific immune
CC responses, e.g. in vaccinated subjects, to prevent an autoimmune reaction
CC to cancer immunotherapy (i.e. to prevent attack on normal but CEA-
CC expressing cells). (Ia) are more active than native sequence (I) and
CC generate a highly specific and systemic anti-CEA response. Cytotoxic T
CC cells generated recognize both (Ia) and native CEA epitopes. The present

```
CC   sequence represents a specifically claimed example of (Ia)
XX
SQ   Sequence 9 AA;

  Query Match           100.0%;  Score 45;  DB 2;  Length 9;
  Best Local Similarity  100.0%;  Pred. No. 1.8e+06;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 YLSGANINL 9
              |||||||||
Db          1 YLSGANINL 9


RESULT 2
AAW39723
ID   AAW39723 standard; peptide; 9 AA.
XX
AC   AAW39723;
XX
DT   11-JUN-1998  (first entry)
XX
DE   Human carcina-embryonic antigen (CEA) peptide (pos. 571-579).
XX
KW   T cell epitope; immune response; human leukocyte antigen; HLA Class I;
KW   vaccine; immunogenic; major histocompatablilty complex; MHC; B cell;
KW   disease; anti-tumour; anti-viral.
XX
OS   Homo sapiens.
XX
PN   WO9741440-A1.
XX
PD   06-NOV-1997.
XX
PF   28-APR-1997;   97WO-NL000229.
XX
PR   26-APR-1996;   96EP-00201145.
PR   23-DEC-1996;   96EP-00203670.
XX
PA   (UYLE-) RIJKSUNIV LEIDEN.
PA   (SCIS-) SCI SEED CAPITAL INVESTMENTS BV.
XX
PI   Van Der Burg SH,  Kast WM,  Toes REM,  Offringa R,  Melief CJM;
XX
DR   WPI; 1997-549891/50.
XX
PT   Method of selecting T cell peptide epitope(s) - by measuring the
PT   stability of HLA class I-peptide complexes on intact B cells.
XX
PS   Example 3; Page 85; 109pp; English.
XX
CC   Peptides AAW39430-W39734 are used in a novel method for the selection of
CC   immunogenic T-cell peptide epitopes present in polypeptide antigens. The
CC   method involves the identification of peptide sequences capable of
CC   binding to an HLA (human leukocyte antigen) class I molecule and
CC   measuring the binding of this epitope peptide to the HLA class I peptide.
CC   The stability of binding of the peptide and MHC (major histocompatablilty
CC   complex) class I molecule is measured on intact human B cells carrying
CC   the MHC molecule at their cell surfaces. The method can be used to select
CC   peptide epitopes for generating vaccines against a disease associated
CC   with the polypeptide, e.g. cancers or AIDS. The peptide epitopes are
CC   especially T-cell peptide epitopes with strong anti-tumour and anti-viral
CC   immune responses. Peptide AAW39723 is derived from the human carcino-
CC   embryonic antigen (CEA) and has the ability to bind to the human MHC
CC   Class I allele HLA-A2.1
XX
SQ   Sequence 9 AA;

  Query Match           95.6%;  Score 43;  DB 2;  Length 9;
  Best Local Similarity  88.9%;  Pred. No. 1.8e+06;
  Matches    8;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 YLSGANINL 9
              ||||||:||
Db          1 YLSGANLNL 9


RESULT 3
AAW70045
ID   AAW70045 standard; peptide; 9 AA.
XX
AC   AAW70045;
XX
DT   22-OCT-1998  (first entry)
XX
DE   CEA derived HLA-A2.1 binding peptide 2 (residues 605-613).
XX
KW   Cytotoxic T lymphocyte; CTL; major histocompatibility complex; MHC;
KW   human leukocyte antigen; HLA; tumour associated antigen; cancer;
KW   antigen presenting cell; APC; immunogenic peptide; immune disorder;
KW   viral infection; AIDS; hepatitis; bacterial infection; malaria; CEA;
KW   fungal infection; tuberculosis; melanoma; carcinoembryonic antigen.
XX
OS   Synthetic.
OS   Homo sapiens.
XX
PN   WO9833888-A1.
XX
PD   06-AUG-1998.
XX
PF   30-JAN-1998;   98WO-US001959.
XX
PR   31-JAN-1997;   97US-0036696P.
XX
PA   (EPIM-) EPIMMUNE INC.
XX
PI   Tsai V,  Southwood S,  Sidney J,  Sette A,  Celis E;
XX
DR   WPI; 1998-437445/37.
XX
PT   Production of antigen-specific cytotoxic T cells - by incubating
PT   immunogenic peptide(s) from antigen that binds class I major
PT   histocompatibility complex molecules with pre-treated antigen presenting
PT   cells.
XX
PS   Example 6; Page 75; 104pp; English.
XX
CC   Sequences shown in AAW70044 to AAW70052 represent peptides derived from
CC   carcinoembryonic antigen (CEA). The peptides can bind to a human
CC   leukocyte antigen (HLA), HLA-A2.1 and are used to exemplify the method of
CC   invention of producing antigen-specific cytotoxic T cells (CTLs) in
CC   vitro. The method comprises contacting immunogenic peptides from an
CC   antigen that binds class I major histocompatibility complex (MHC)
CC   molecules with antigen presenting cells (APCs) pretreated with
CC   pretreatment growth factors, and incubating the APCs with purified CD8
CC   cells in the presence of at least 2 incubation growth factors, thereby
CC   producing antigen-specific CTLs. A method for specifically killing target
CC   cells in a human patient is also provided which comprises obtaining a
CC   fluid sample containing CTLs from a patient, contacting the cytotoxic T
CC   cells with APCs pretreated with pre-treatment growth factors, where the
CC   APCs comprise class I MHC molecules. The pretreated APCs are incubated
CC   with the cytotoxic growth factors, thereby producing activated CTLs which
CC   are contacted with a carrier to form a composition. The composition can
CC   then be administered to the patient. The activated CTLs can be used for
CC   treating cancers, immune disorders, viral infections, AIDS, hepatitis,
CC   bacterial infection, fungal infection, malaria or tuberculosis
XX
SQ   Sequence 9 AA;

  Query Match           95.6%;  Score 43;  DB 2;  Length 9;
  Best Local Similarity  88.9%;  Pred. No. 1.8e+06;
  Matches    8;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 YLSGANINL 9
              ||||||:||
Db          1 YLSGANLNL 9
```

RESULT 4
AAW77134
ID AAW77134 standard; peptide; 9 AA.
XX
AC AAW77134;
XX
DT 16-NOV-1998 (first entry)
XX
DE CEA synthetic peptide epitope 1.
XX
KW Tyrosinase; tyrosinase cytotoxic lymphocyte response;
KW cytotoxic T lymphocyte; cysteine-depleted; melanoma.
XX
OS Synthetic.
XX
PN WO9833810-A2.
XX
PD 06-AUG-1998.
XX
PF 29-JAN-1998; 98WO-US001592.
XX
PR 30-JAN-1997; 97US-0037781P.
XX
PA (UYVI-) UNIV VIRGINIA PATENT FOUND.
XX
PI Slingluff CL, Hunt DF, Engelhard VH, Kittlesen D;
XX
DR WPI; 1998-437388/37.
XX
PT Disease specific immunogen - comprises disease specific cytotoxic T
PT lymphocyte epitope used to elicit melanoma specific CTL response.
XX
PS Disclosure; Page 27; 93pp; English.
XX
CC The peptide epitope AAW77119-W77138 were created for human tumour-
CC specific cytotoxic T lymphocyte response. These peptides are are cysteine
CC - depleted mutants of a native disease-specific CTL epitope. The cysteine
CC - depleted CTL epitopes elicit a stronger or more specific CTL response
CC than the native epitope. The epitopes can be used in a disease-specific
CC immunogen to protect a mammal against disease in particular melanomas.
CC The peptides may also be used to screen a sample for the presence of an
CC antigen with the same epitope, or with a different cross-reactive epitope
XX
SQ Sequence 9 AA;

Query Match 95.6%; Score 43; DB 2; Length 9;
Best Local Similarity 88.9%; Pred. No. 1.8e+06;
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 YLSGANINL 9
||||||:||
Db 1 YLSGANLNL 9

RESULT 5
AAY47655
ID AAY47655 standard; peptide; 9 AA.
XX
AC AAY47655;
XX
DT 01-DEC-1999 (first entry)
XX
DE Immunogenic peptide having a human leukocyte antigen binding motif #2266.
XX
KW Human leukocyte antigen; binding; immunogenic; glycoprotein; MHC; HLA;
KW immune response; T cell activation; major histocompatibility complex;
KW cytotoxic T lymphocyte; CTL; tumour rejection; viral infection; cancer;
KW prostate cancer; hepatitis B; hepatitis C; AIDS; renal carcinoma;
KW vaccine; immunisation.
XX
OS Synthetic.
OS Homo sapiens.
XX
PN WO9945954-A1.
XX
PD 16-SEP-1999.
XX
PF 13-MAR-1998; 98WO-US005039.
XX
PR 13-MAR-1998; 98WO-US005039.
XX
PA (EPIM-) EPIMMUNE INC.
XX
PI Sette A, Kubo RT, Sidney J, Celis E, Grey HM, Southwood S;
XX
DR WPI; 1999-551214/46.
XX
PT New immunogenic peptides with HLA binding motif, useful in treatment and
PT diagnosis of cancers and viral diseases.
XX
PS Claim 1; Page 118; 150pp; English.
XX
CC AAY45390 to AAY48214 represent specifically claimed immunogenic peptides
CC having a human major histocompatibility complex (MHC) Class I (also known
CC as human leukocyte antigen (HLA)) binding motif. The immunogenic peptides
CC can bind to a specific HLA allele (i.e. HLA-A subtypes HLA-A2.1, A1, A3.2
CC or A24.1 or HLA-B or C) and induce a cytotoxic T cell response against
CC the antigen from which the peptide is derived. Cytotoxic T lymphocytes
CC (CTLs) which destroy antigen-bearing cells are normally induced by an
CC antigen in the form of a peptide fragment bound to a HLA molecule, rather
CC than the intact foreign antigen itself, and are particularly important in
CC tumour rejection and in fighting viral infections. The peptides are
CC therefore useful therapeutically to treat or prevent viral infections and
CC cancers in mammals (especially humans) e.g. prostate cancer, hepatitis B
CC and C, AIDS, and renal carcinoma. They can be administered as vaccines to
CC elicit an immune response in individuals susceptible or otherwise at risk
CC of viral infection or cancer, or used to treat chronic or acute
CC conditions. They are also useful diagnostically, and can be used to
CC induce a cytotoxic T cell response, by contacting a cytotoxic T cell with
CC the peptide e.g. to produce CTLs ex vivo for infusion back into a
CC patient. The polynucleotides encoding the immunogenic peptides are also
CC useful therapeutically and for immunisation as above
XX
SQ Sequence 9 AA;

Query Match 95.6%; Score 43; DB 2; Length 9;
Best Local Similarity 88.9%; Pred. No. 1.8e+06;
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 YLSGANINL 9
||||||:||
Db 1 YLSGANLNL 9

RESULT 6
AAY09525
ID AAY09525 standard; peptide; 9 AA.
XX
AC AAY09525;
XX
DT 20-JUL-1999 (first entry)
XX
DE Carcinoembryonic antigen peptide agonist CAP-1.
XX
KW Carcinoembryonic antigen; CEA; human; agonist; antagonist;
KW immune response; carcinoma; gastrointestinal; breast; pancreatic;
KW bladder; ovarian; lung; prostatic; T cell proliferation; cancer;
KW adoptive transfer therapy; autoimmune reaction; immunotherapy.
XX
OS Homo sapiens.
OS Synthetic.
XX
PN WO9919478-A1.
XX
PD 22-APR-1999.

XX
PF 22-SEP-1998; 98WO-US019794.
XX
PR 10-OCT-1997; 97US-0061589P.
XX
PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
XX
PI Schlom J, Barzaga E, Zaremba S;
XX
DR WPI; 1999-326544/27.
XX
PT Peptide agonists and antagonists of carcinoembryonal antigen.
XX
PS Claim 1; Page 53; 72pp; English.
XX
CC The present invention describes peptides (A) that comprise agonists (Ia)
CC or antagonists (Ib) of human carcinoembryonal antigen (CEA). (Ia) are
CC used in vaccines to kill or inhibit carcinoma cells that express CEA or
CC its epitopes, particularly for treating gastrointestinal, breast,
CC pancreatic, bladder, ovarian, lung or prostatic carcinoma. They can also
CC be used to proliferate T cells, e.g. from vaccinated subjects, for use in
CC adoptive transfer therapy. (Ib) are used to inhibit CEA-specific immune
CC responses, e.g. in vaccinated subjects, to prevent an autoimmune reaction
CC to cancer immunotherapy (i.e. to prevent attack on normal but CEA-
CC expressing cells). (Ia) are more active than native sequence (I) and
CC generate a highly specific and systemic anti-CEA response. Cytotoxic T
CC cells generated recognize both (Ia) and native CEA epitopes. The present
CC sequence represents a specifically claimed example of (Ia)
XX
SQ Sequence 9 AA;

Query Match 95.6%; Score 43; DB 2; Length 9;
Best Local Similarity 88.9%; Pred. No. 1.8e+06;
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 YLSGANINL 9
||||||:||
Db 1 YLSGANLNL 9

RESULT 7
AAB13749
ID AAB13749 standard; peptide; 9 AA.
XX
AC AAB13749;
XX
DT 02-FEB-2001 (first entry)
XX
DE Peptide fragment # 1 from human CEA.
XX
KW Human; T-cell; immune response; antigen; epitope; B7 family molecule;
KW Leukocyte function-associated antigen-3; LFA-3;
KW Intercellular adhesion molecule-1; ICAM-1; vaccine; immunotherapy;
KW colon polyp; Crohn's disease; ulcerative colitis; breast lesion; tumour;
KW CEA.
XX
OS Homo sapiens.
XX
PN WO200034494-A1.
XX
PD 15-JUN-2000.
XX
PF 12-NOV-1999; 99WO-US026866.
XX
PR 09-DEC-1998; 98US-0111582P.
XX
PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
PA (THER-) THERION BIOLOGICS CORP.
XX
PI Schlom J, Hodge J, Panicali D;
XX
DR WPI; 2000-431307/37.
XX
PT Novel recombinant vector useful as immunogens and vaccines for
PT stimulating and enhancing immunological responses to target cells and
PT antigens expresses multiple co-stimulatory molecules such as B7-1, LFA-3,
PT ICAM-1.
XX
PS Claim 18; Page 35; 188pp; English.
XX
CC Costimulatory molecules have important roles in T-cell activation and
CC therefore the immune response. The present invention relates to
CC recombinant vectors which comprise of foreign nucleic acid sequences
CC encoding at least three costimulatory molecules: a B7 family molecule,
CC Leukocyte function-associated antigen-3 (LFA-3, human CD58) and
CC Intercellular adhesion molecule-1 (ICAM-1, CD54) and optionally a foreign
CC gene encoding a target antigen or immunological epitope. The present
CC sequence is one such target antigen used in the present invention. The
CC present sequence is a tumour-associated antigen. The vector of the
CC present invention would be useful for providing an enhanced immune
CC response to the present target antigen. The vector of the present
CC invention may therefore be useful in immunotherapy for treating or
CC preventing diseases caused by viruses, bacteria, protozoans, parasites,
CC premalignant cells and tumour cells. The recombinant vector can be used
CC to treat or prevent preneoplastic or hyperplastic states such as colon
CC polyps, Crohn's disease, ulcerative colitis and breast lesions
XX
SQ Sequence 9 AA;

Query Match 95.6%; Score 43; DB 3; Length 9;
Best Local Similarity 88.9%; Pred. No. 1.8e+06;
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 YLSGANINL 9
||||||:||
Db 1 YLSGANLNL 9

RESULT 8
AAE02673
ID AAE02673 standard; peptide; 9 AA.
XX
AC AAE02673;
XX
DT 06-AUG-2001 (first entry)
XX
DE Human CEA epitopic peptide.
XX
KW Human; cytostatic; antibacterial; antifungal; gene therapy; vaccine;
KW antiviral; tumour; epitope; glycoprotein; hepatitis B virus; HBV;
KW immune response; CTL; cytotoxic T lymphocyte; CEA; HLA;
KW human leucocyte antigen.
XX
OS Homo sapiens.
XX
PN WO200127291-A1.
XX
PD 19-APR-2001.
XX
PF 29-SEP-2000; 2000WO-EP009902.
XX
PR 12-OCT-1999; 99US-0158356P.
XX
PA (INSP) INST PASTEUR.
XX
PI Firat H, Lemonnier F, Langlade-Demoyen P;
XX
DR WPI; 2001-282038/29.
XX
PT New polynucleotide comprising at least one viral, fungal, bacterial, or
PT tumor epitope of an antigen, capable of inducing a cellular response.
XX
PS Example 1; Page 23; 70pp; English.
XX
CC The invention relates to polynucleotide containing at least a part of the
CC coding sequence of the middle glycoprotein of hepatitis B virus (HBV) in

CC which is inserted a DNA sequence coding for an epitope comprising at
CC least one viral, fungal, bacterial, or tumour epitope of an antigen,
CC capable of inducing a cellular response. Nucleic acids and compositions
CC of the invention are useful for inducing in vivo a CTL (cytotoxic T
CC lymphocyte) response against several epitopes of one or more, bacterial,
CC viral, fungal, or tumour antigens. A composition of the invention
CC produces an immune response against HIV antigen and are used in the
CC production of vaccines. The polynucleotides of the invention are also
CC used in gene therapy. The present sequence is human CEA epitopic peptide.
CC This peptide elicits strong HLA (human leucocyte antigen)-A2.1-restricted
CC CTL response in mice
XX
SQ Sequence 9 AA;

Query Match 95.6%; Score 43; DB 4; Length 9;
Best Local Similarity 88.9%; Pred. No. 1.8e+06;
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 YLSGANINL 9
||||||:||
Db 1 YLSGANLNL 9

RESULT 9
AAE00463
ID AAE00463 standard; peptide; 9 AA.
XX
AC AAE00463;
XX
DT 19-JUN-2001 (first entry)
XX
DE Human tumour CEA epitopic peptide.
XX
KW Human; tumour epitope; cytostatic; immunostimulant; gene therapy;
KW middle glycoprotein; Hepatitis B virus; HBV; cytotoxic response;
KW immune response; cytotoxic T lymphocyte; CTL; CEA; HLA;
KW human leucocyte antigen.
XX
OS Homo sapiens.
XX
PN WO200123577-A2.
XX
PD 05-APR-2001.
XX
PF 29-SEP-2000; 2000WO-EP009900.
XX
PR 30-SEP-1999; 99US-0156945P.
XX
PA (INSP) INST PASTEUR.
XX
PI Firat H, Lemonnier F, Langlade-Demoyen P, Michel M, Suhrbier AA;
XX
DR WPI; 2001-266164/27.
XX
PT Novel polynucleotide having DNA sequence encoding tumor antigen epitope
PT inserted in part of coding sequence of middle glycoprotein of hepatitis B
PT virus, used to induce immune response against tumor-specific antigen.
XX
PS Example 1; Page 13; 36pp; English.
XX
CC The present invention relates to an isolated or purified polynucleotide
CC containing a DNA sequence coding for at least one tumour epitope of a
CC tumour antigen inserted into part of the coding sequence of the middle
CC glycoprotein of the Hepatitis B virus (HBV). The polynucleotide is useful
CC for optionally evaluating cytotoxic responses in the individual's
CC lymphocyte population. It induces an immune response against at least one
CC tumour specific antigen or tissue specific antigen. The vector comprising
CC the polynucleotide induces in vivo, cellular and/or humoral immune
CC response. The composition comprising the polynucleotide induces in vivo,
CC cytotoxic T lymphocyte (CTL) against one or more antigens or epitopes
CC present on the hybrid protein. The polynucleotide is also useful in gene
CC therapy. The present sequence is a human tumour CEA epitopic peptide.
CC This peptide elicits strong HLA (human leucocyte antigen)-A2.1-restricted
CC CTL response in mice
XX
SQ Sequence 9 AA;

Query Match 95.6%; Score 43; DB 4; Length 9;
Best Local Similarity 88.9%; Pred. No. 1.8e+06;
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 YLSGANINL 9
||||||:||
Db 1 YLSGANLNL 9

RESULT 10
AAE05123
ID AAE05123 standard; peptide; 9 AA.
XX
AC AAE05123;
XX
DT 18-SEP-2001 (first entry)
XX
DE Carcinoembryonic antigen (CEA) peptide, CAP-1.
XX
KW Tumour-associated antigen; TAA; cytostatic; vaccine; gene therapy;
KW immune response; tetanus toxoid; TT; diphtheria toxoid; DT; prophylactic;
KW cancer; therapeutic; carcinoembryonic antigen; CEA.
XX
OS Unidentified.
XX
PN WO200149317-A2.
XX
PD 12-JUL-2001.
XX
PF 05-JAN-2001; 2001WO-CA000005.
XX
PR 05-JAN-2000; 2000US-0174587P.
XX
PA (AVET) AVENTIS PASTEUR LTD.
XX
PI Emtage P, Barber BH, Sambhara S, Sia CDY;
XX
DR WPI; 2001-441790/47.
XX
PT Enhancing immune response to antigen such as tumor antigen for treating
PT cancer in an animal involves administering an inducing agent to the
PT animal followed by administering inducing agent-antigen mixture.
XX
PS Example 2; Page 31; 62pp; English.
XX
CC The invention relates to a method of enhancing an immune response against
CC tumour-associated antigens (TAAs), such as GP100 and carcinoembryonic
CC antigen (CEA) in an animal. The method involves priming of the animal
CC with an inducing agent such as tetanus toxoid (TT) or diphtheria toxoid
CC (DT), subsequently followed by administration of an inducing agent-
CC antigen mixture. The method provides the enhancement or augmentation of
CC the immune response to the antigen and/or improves a vaccination protocol
CC by allowing use of less antigen. The immunisation of the animal with
CC tumour-associated antigen is useful for the prophylactic or therapeutic
CC treatment of cancer. The present sequence is carcinoembryonic antigen
CC (CEA) peptide fragment related to the invention
XX
SQ Sequence 9 AA;

Query Match 95.6%; Score 43; DB 4; Length 9;
Best Local Similarity 88.9%; Pred. No. 1.8e+06;
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 YLSGANINL 9
||||||:||
Db 1 YLSGANLNL 9

RESULT 11

```
AAB82776
ID   AAB82776 standard; protein; 9 AA.
XX
AC   AAB82776;
XX
DT   29-OCT-2001  (first entry)
XX
DE   Carcinoembryonic antigen peptide.
XX
KW   Telomerase reverse transcriptase; hTRT; human; cytotoxic T lymphocyte;
KW   major histocompatibility complex; cancer; tumour;
KW   human leucocyte antigen; HLA-A2.1; vaccine; carcinoembryonic antigen.
XX
OS   Homo sapiens.
XX
PN   WO200160391-A1.
XX
PD   23-AUG-2001.
XX
PF   15-FEB-2001; 2001WO-US005143.
XX
PR   15-FEB-2000; 2000US-0182685P.
PR   15-FEB-2001; 2001US-00182685.
XX
PA   (REGC ) UNIV CALIFORNIA.
XX
PI   Zanetti M;
XX
DR   WPI; 2001-536552/59.
XX
PT   Vaccine for initiating and enhancing a cytotoxic T lymphocyte response,
PT   for treating cancers or tumors or for inducing immune response against
PT   tumors, comprises a telomerase reverse transcriptase peptide.
XX
PS   Example 1; Page 12; 52pp; English.
XX
CC   The present sequence is that of a carcinoembryonic antigen peptide
CC   comprising amino acid residues 571-579. The peptide was used as a
CC   reference peptide in comparison with human telomerase reverse
CC   transcriptase (hTRT) HLA-A2.1+ restricted peptide p540 (see AAB82772) in
CC   a HLA-A2.1 binding/stabilisation assay. The induction of CTL responses in
CC   vitro and in vivo, and the susceptibility to lysis of tumour cells of
CC   various origins by hTRT CTL suggest that hTRT could serve as a universal
CC   cancer vaccine for humans. A claimed universal vaccine for treating
CC   tumours of any origin comprises at least 1 hTRT peptide. The peptide is 7
CC   -15 amino acid residues in length and may be modified to enhance binding
CC   to the major histocompatibility complex. Also claimed is a method for
CC   inducing and enhancing a CTL response against cancer cells, involving
CC   harvesting blood leucocytes, pulsing with hTRT, and contacting cancer
CC   cells with the pulsed leucocytes. A method for targeting CTL to tumour
CC   cells is also claimed, and involves administering a hTRT peptide to a
CC   mammal, especially a cancer patient
XX
SQ   Sequence 9 AA;

  Query Match             95.6%;  Score 43;  DB 4;  Length 9;
  Best Local Similarity   88.9%;  Pred. No. 1.8e+06;
  Matches    8;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 YLSGANINL 9
              ||||||:||
Db          1 YLSGANLNL 9


RESULT 12
ABG79073
ID   ABG79073 standard; peptide; 9 AA.
XX
AC   ABG79073;
XX
DT   15-NOV-2002  (first entry)
XX
DE   Human CEA class I HLA widely expressed antigen peptide #1.
XX
KW   Cell penetrating peptide; cancer; tumour; melanoma; thymoma; antigen;
KW   lymphoma; sarcoma; lung cancer; non-Hodgkin's lymphoma; leukaemia;
KW   Hodgkin's lymphoma; uterine cancer; cervical cancer; bladder cancer;
KW   kidney cancer; adenocarcinoma; breast cancer; prostate cancer;
KW   ovarian cancer; pancreatic cancer; epitope; vaccine; dendritic cell;
KW   tumour infiltrating lymphocyte; TIL; human leukocyte antigen; HLA;
KW   cytostatic; human.
XX
OS   Homo sapiens.
XX
PN   WO200264057-A2.
XX
PD   22-AUG-2002.
XX
PF   15-FEB-2002; 2002WO-US005212.
XX
PR   15-FEB-2001; 2001US-0268687P.
XX
PA   (BAYU ) BAYLOR COLLEGE MEDICINE.
XX
PI   Wang R;
XX
DR   WPI; 2002-627577/67.
XX
PT   Novel composition for treating a disease in an animal, comprises an
PT   immune effector cell and cell penetrating peptide associated with an
PT   antigen or antibody.
XX
PS   Disclosure; Page 17; 61pp; English.
XX
CC   The invention relates to a composition (I) comprising an immune effector
CC   cell and a cell penetrating peptide (CPP) associated with an antigen or
CC   antibody. Also included are (1) a vaccine comprising (I), CPP associated
CC   with an antigen, and a pharmaceutically acceptable carrier and (2)
CC   preparing a composition for a disease, by providing (I) and CPP
CC   associated with an antigen for disease, and introducing the antigen-
CC   associated CPP to (I), where antigen enters into the cell. The antigens
CC   are, for example, tumour antigen derived epitopes recognised by tumour
CC   infiltrating lymphocytes (TIL) of HLA (human leukocyte antigen) class I
CC   or II. The composition is useful for enhancing immunity in an animal to a
CC   disease, by administering a mature dendritic cell comprising CPP
CC   associated with an antigen to disease, to the animal, such that following
CC   the administration, animal is protected from disease, where the animal
CC   comprises both CD4+ and CD8+ T cells. It is also useful for treating a
CC   disease (e.g. cancer, tumour, melanoma, thymoma, lymphoma, sarcoma, lung
CC   cancer, non-Hodgkin's lymphoma, leukaemia, Hodgkin's lymphoma, uterine
CC   cancer, cervical cancer, bladder cancer, kidney cancer, adenocarcinoma,
CC   breast cancer, prostate cancer, ovarian cancer and pancreatic cancer).
CC   The animal is further subjected to a cancer treatment including surgery,
CC   radiation, chemotherapy or gene therapy. The administration of (I),
CC   preferably dendritic cell is prior to, subsequent to or concurrent with,
CC   the cancer treatment. The present sequence is a tumour antigen derived
CC   epitope for inclusion in the composition of the invention
XX
SQ   Sequence 9 AA;

  Query Match             95.6%;  Score 43;  DB 5;  Length 9;
  Best Local Similarity   88.9%;  Pred. No. 1.8e+06;
  Matches    8;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 YLSGANINL 9
              ||||||:||
Db          1 YLSGANLNL 9


RESULT 13
AAE26805
ID   AAE26805 standard; peptide; 9 AA.
XX
AC   AAE26805;
XX
DT   13-DEC-2002  (first entry)
```

XX
DE Human HLA-A2.1 restricted CEA (carcinoembryonic Ag) peptide epitope.
XX
KW Human; cancer; breast cancer; ovarian cancer; melanoma; cell therapy;
KW epitope; human leucocyte antigen; HLA-A2.1.
XX
OS Homo sapiens.
XX
PN WO200265992-A2.
XX
PD 29-AUG-2002.
XX
PF 19-FEB-2002; 2002WO-US005748.
XX
PR 20-FEB-2001; 2001US-0270252P.
XX
PA (ORTH) ORTHO-MCNEIL PHARM INC.
XX
PI Degraw J, Moriarty A, Leturcq DJ, Jackson MR, Peterson PA;
PI Heiskala M;
XX
DR WPI; 2002-667033/71.
XX
PT Treating a subject with cancer comprises combining the CD+8 cells, which
PT are stimulated with non-naturally occurring antigen-presenting cell line,
PT with adherent blood monocytes and inoculating the subject with CD8+
PT suspension.
XX
PS Example 2; Page 93; 99pp; English.
XX
CC The invention relates to a method of treating a subject with cancer. The
CC method involves combining the CD+8 cells, which are stimulated with non
CC naturally occurring antigen-presenting cell (nnAPC) line, with adherent
CC blood monocytes and inoculating the subject with CD8+ suspension. The
CC method is useful for treating cancer e.g. ovarian cancer, breast cancer
CC and melanoma etc. It is also useful in cell therapy. The present sequence
CC is human leukocyte antigen A2 (HLA-A2).1 restricted peptide epitope used
CC to treat breast and ovarian cancer
XX
SQ Sequence 9 AA;

Query Match 95.6%; Score 43; DB 5; Length 9;
Best Local Similarity 88.9%; Pred. No. 1.8e+06;
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 YLSGANINL 9
||||||:||
Db 1 YLSGANLNL 9

RESULT 14
AAU95893
ID AAU95893 standard; peptide; 9 AA.
XX
AC AAU95893;
XX
DT 02-JUL-2002 (first entry)
XX
DE Immunogenic peptide with (HLA)-A2.1 binding site #106.
XX
KW HLA-A2.1 binding peptide; cytostatic; virucide; anti-HIV; hepatotropic;
KW human immunodeficiency virus; antiinflammatory; antibacterial; vaccine;
KW protozoacide; immunosuppressant; immunogenic peptide; T cell activation;
KW human leucocyte antigen binding site; cytotoxic T cell response;
KW viral infection; hepatitis; Epstein-Barr virus; papilloma virus;
KW human immunodeficiency virus; HIV; Kaposi sarcoma; Lassa fever virus;
KW cytomegalovirus; tumour; prostate cancer; renal carcinoma; lymphona;
KW prostate-specific antigen; p53; carcino-embryonal antigen;
KW melanoma antigen; Mycobacterium tuberculosis; protozoa;
KW trypanosome surface antigen; condlyloma acuminatum.
XX
OS Unidentified.
XX
PN WO200220616-A1.
XX
PD 14-MAR-2002.
XX
PF 01-SEP-2000; 2000WO-US024102.
XX
PR 01-SEP-2000; 2000WO-US024102.
XX
PA (EPIM-) EPIMMUNE INC.
XX
PI Grey HM, Sette A, Sidney J, Southwood S;
XX
DR WPI; 2002-351766/38.
XX
PT Immunogenic peptide with human leucocyte antigen-A2.1 binding site,
PT useful for treating e.g. viral infection or tumors.
XX
PS Claim 1; Page 27; 35pp; English.
XX
CC The invention describes a composition comprising an immunogenic peptide
CC having a human leucocyte antigen (HLA)-A2.1 binding site. The peptides
CC bind specifically to HLA-A2.1, to cause T cell activation and thus a
CC cytotoxic T cell response. The peptides and the nucleic acids that
CC encodes them, are used, in vivo or ex vivo, for treatment of viral
CC infections (hepatitis B or C; Epstein-Barr; human immune deficiency;
CC Kaposi sarcoma; human papilloma; Lassa fever or cytomegaloviruses);
CC tumours including prostate cancer, renal carcinoma and lymphona (where
CC directed to prostate-specific antigen, p53, carcino-embryonal antigen,
CC Her2/neu or melanoma antigens); infection by Mycobacterium tuberculosis
CC or protozoa (directed to trypanosome surface antigen); and condlyloma
CC acuminatum. The peptides are suitable for use in peptide-based vaccines.
CC This sequence represents an immunogenic peptide with the human leucocyte
CC antigen (HLA)-A2.1 binding site, described in the invention
XX
SQ Sequence 9 AA;

Query Match 95.6%; Score 43; DB 5; Length 9;
Best Local Similarity 88.9%; Pred. No. 1.8e+06;
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 YLSGANINL 9
||||||:||
Db 1 YLSGANLNL 9

RESULT 15
AAE19088
ID AAE19088 standard; peptide; 9 AA.
XX
AC AAE19088;
XX
DT 21-MAY-2002 (first entry)
XX
DE HLA-A24 restricted target antigen CEA immunological epitope #2.
XX
KW Human leukocyte antigen; HLA; pharmaceutical composition; target antigen;
KW immunological epitope; replication-defective virus; RDV; immune response;
KW chemotherapy; granulocyte-monocyte-colony stimulating factor; cytostatic;
KW GM-CSF; MHC; major histocompatibility complex; tumour; head; pancreatic;
KW neck; breast; prostate; colorectal; melanoma; myeloidysplastic syndrome;
KW metastatic breast skin lesion; corticosteroid therapy; erythropoietin;
KW cytopenia; neutropenia; vaccine; immunostimulant.
XX
OS Homo sapiens.
XX
PN WO200195919-A2.
XX
PD 20-DEC-2001.
XX
PF 15-JUN-2001; 2001WO-US019201.
XX
PR 15-JUN-2000; 2000US-0211717P.
XX

PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
PA (THER-) THERION BIOLOGICS CORP.
XX
PI Schlom J, Greiner JW, Kass E, Panicali D;
XX
DR WPI; 2002-205852/26.
XX
PT Composition for enhancing immune responses, particularly anti-tumor
PT responses and treating neutropenia, cytopenia, comprises replication-
PT defective virus encoding granulocyte-monocyte-colony stimulating factor.
XX
PS Claim 9; Page 15; 118pp; English.
XX
CC The present invention relates to a pharmaceutical composition comprising
CC a replication-defective virus (RDV) encoding granulocyte-monocyte-colony
CC stimulating factor (GM-CSF). The invention is useful for enhancing cell-
CC mediated or humoral immune response in an individual, by enhancing
CC migration of APC expressing CD11c^+/I-Ab^+, major histocompatibility
CC complex (MHC) class II, at an injection site, regional lymph node at a
CC tumour site, APC proliferation or function, CD4^+T or CD8^+T cell
CC activation, interleukin (IL)-2, interferon (IFN)-gamma or tumour necrosis
CC factor (TNF)-alpha production or their combinations. The composition
CC enhances an antigen-specific T-cell response in an individual to a target
CC antigen or its immunological epitope and an anti-tumour response in an
CC individual with a head tumour, neck, breast, pancreatic, prostate,
CC colorectal or metastatic tumour or melanoma, or metastatic breast skin
CC lesion. The invention is further useful for treating neutropenia
CC resulting from chemotherapy, corticosteroid therapy, irradiation or an
CC infection, by raising the neutrophil count to normal levels and for
CC treating cytopenias in patients with myeloidysplastic syndrome in
CC combination with erythropoietin, by increasing neutrophil count and
CC erythroid precursors. The composition enhances immune response to
CC vaccines such as DPT, Td, DtaP, Hib, DtaP-Hib, MMR, Hepatitis A,
CC hepatitis B, lyme's disease, influenza, tetravalent meningococcal
CC polysaccharide, pneumococcal polysaccharide, anthrax, cholera, plague,
CC yellow fever and Bacillus Calmette-Guerin vaccine. The present sequence
CC is human leukocyte antigen (HLA)-restricted target tumour antigen
CC immunological epitope
XX
SQ Sequence 9 AA;

Query Match 95.6%; Score 43; DB 5; Length 9;
Best Local Similarity 88.9%; Pred. No. 1.8e+06;
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 YLSGANINL 9
              |||||:||
Db          1 YLSGANLNL 9
```

Search completed: May 17, 2005, 06:17:51
Job time : 67 secs

GenCore version 5.1.6

OM protein - nucleic search, using frame_plus_p2n model

Run on: May 17, 2005, 16:29:39 ; Search time 326.5 Seconds
(without alignments)
163.178 Million cell updates/sec

Title: US-10-725-373-4
Perfect score: 45
Sequence: 1 YLSGANINL 9

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/US10725373/runat_17052005_071020_16177/app_query.fasta_1.796
-DB=N_Geneseq_16Dec04 -QFMT=fastap -SUFFIX=rng -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10725373_@CGN_1_1_1241_@runat_17052005_071020_16177 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N_Geneseq_16Dec04:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002as:*
7: geneseqn2002bs:*
8: geneseqn2003as:*
9: geneseqn2003bs:*
10: geneseqn2003cs:*
11: geneseqn2003ds:*
12: geneseqn2004as:*
13: geneseqn2004bs:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	43	95.6	27	2	AAX56260	Aax56260 Carcinoem
	2	43	95.6	30	12	ADL46174	Adl46174 Human CAP
	3	43	95.6	64	12	ADL46175	Adl46175 Human imm
c	4	43	95.6	80	2	AAV57948	Aav57948 708 vkcea
c	5	43	95.6	80	2	AAV81101	Aav81101 Vaccine 2
	6	43	95.6	155	4	AAI29234	Aai29234 Colon tum
	7	43	95.6	155	8	ABZ33420	Abz33420 Human col
	8	43	95.6	256	4	AAS57750	Aas57750 cDNA #426
c	9	43	95.6	340	6	ABV88334	Abv88334 Human col
c	10	43	95.6	407	4	AAS57366	Aas57366 cDNA #42
c	11	43	95.6	409	4	AAS57425	Aas57425 cDNA #101
	12	43	95.6	409	6	ABV86774	Abv86774 Human col
c	13	43	95.6	409	6	ABV87551	Abv87551 Human col
	14	43	95.6	409	6	ABV89100	Abv89100 Human col
	15	43	95.6	409	6	ABV87855	Abv87855 Human col
	16	43	95.6	409	6	ABK39290	Abk39290 DNA encod
	17	43	95.6	409	6	ABK39002	Abk39002 cDNA enco
c	18	43	95.6	409	6	ABK39424	Abk39424 DNA encod
	19	43	95.6	409	6	ABK45946	Abk45946 cDNA enco
	20	43	95.6	409	6	ABK45289	Abk45289 cDNA enco
c	21	43	95.6	409	6	ABK27782	Abk27782 Human col
	22	43	95.6	409	8	ACA11619	Aca11619 Human lun
	23	43	95.6	409	8	ACA11331	Aca11331 Human lun
c	24	43	95.6	409	8	ACA11753	Aca11753 Human lun
c	25	43	95.6	409	8	ACA02939	Aca02939 Lung canc
	26	43	95.6	409	8	ACA02805	Aca02805 Lung canc
	27	43	95.6	409	8	ACA02517	Aca02517 Lung canc
	28	43	95.6	409	10	ADH46559	Adh46559 Human lun
	29	43	95.6	409	10	ADH46847	Adh46847 Human lun
c	30	43	95.6	409	10	ADH46981	Adh46981 Human lun
	31	43	95.6	409	13	ADJ20766	Adj20766 Human lun
	32	43	95.6	409	13	ADJ20478	Adj20478 Human lun
c	33	43	95.6	409	13	ADJ20900	Adj20900 Human lun
c	34	43	95.6	410	4	AAS57430	Aas57430 cDNA #106
	35	43	95.6	410	4	AAS57434	Aas57434 cDNA #110
c	36	43	95.6	410	6	ABV88264	Abv88264 Human col
	37	43	95.6	412	3	AAA77726	Aaa77726 cDNA enco
	38	43	95.6	412	4	AAI28464	Aai28464 Colon tum
	39	43	95.6	412	8	ABZ32650	Abz32650 Human col
	40	43	95.6	413	4	AAS57627	Aas57627 cDNA #303
	41	43	95.6	415	5	AAF68500	Aaf68500 Human lun
	42	43	95.6	415	6	ABK38411	Abk38411 cDNA enco
	43	43	95.6	415	8	ACA10740	Aca10740 Human lun
	44	43	95.6	415	8	ABX99691	Abx99691 Lung canc
	45	43	95.6	415	10	ADH45954	Adh45954 Human lun

ALIGNMENTS

RESULT 1
AAX56260
ID AAX56260 standard; DNA; 27 BP.
XX
AC AAX56260;
XX
DT 20-JUL-1999 (first entry)
XX
DE Carcinoembryonic antigen peptide agonist CAP-1 encoding DNA SEQ ID NO:6.
XX
KW Carcinoembryonic antigen; CEA; human; agonist; antagonist;
KW immune response; carcinoma; gastrointestinal; breast; pancreatic;
KW bladder; ovarian; lung; prostatic; T cell proliferation; cancer;
KW adoptive transfer therapy; autoimmune reaction; immunotherapy; ss.
XX
OS Homo sapiens.
OS Synthetic.
XX
PN WO9919478-A1.
XX
PD 22-APR-1999.
XX
PF 22-SEP-1998; 98WO-US019794.
XX
PR 10-OCT-1997; 97US-0061589P.
XX
PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
XX
PI Schlom J, Barzaga E, Zaremba S;

XX
DR WPI; 1999-326544/27.
XX
PT Peptide agonists and antagonists of carcinoembryonal antigen.
XX
PS Disclosure; Page 19; 72pp; English.
XX
CC The present invention describes peptides (A) that comprise agonists (Ia)
CC or antagonists (Ib) of human carcinoembryonal antigen (CEA). (Ia) are
CC used in vaccines to kill or inhibit carcinoma cells that express CEA or
CC its epitopes, particularly for treating gastrointestinal, breast,
CC pancreatic, bladder, ovarian, lung or prostatic carcinoma. They can also
CC be used to proliferate T cells, e.g. from vaccinated subjects, for use in
CC adoptive transfer therapy. (Ib) are used to inhibit CEA-specific immune
CC responses, e.g. in vaccinated subjects, to prevent an autoimmune reaction
CC to cancer immunotherapy (i.e. to prevent attack on normal but CEA-
CC expressing cells). (Ia) are more active than native sequence (I) and
CC generate a highly specific and systemic anti-CEA response. Cytotoxic T
CC cells generated recognize both (Ia) and native CEA epitopes. The present
CC sequence encodes a specifically claimed example of (Ia)
XX
SQ Sequence 27 BP; 6 A; 10 C; 5 G; 6 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	0.0935	Length:	27
Score:	43.00	Matches:	8
Percent Similarity:	100.00%	Conservative:	1
Best Local Similarity:	88.89%	Mismatches:	0
Query Match:	95.56%	Indels:	0
DB:	2	Gaps:	0

US-10-725-373-4 (1-9) x AAX56260 (1-27)

```
Qy          1 TyrLeuSerGlyAlaAsnIleAsnLeu 9
              |||||||||||||||||:::||||||
Db          1 TACCTTTCGGGAGCGAACCTCAACCTC 27
```

RESULT 2
ADL46174
ID ADL46174 standard; DNA; 30 BP.
XX
AC ADL46174;
XX
DT 17-JUN-2004 (first entry)
XX
DE Human CAP-1 tumour antigen fragment DNA, SEQ ID NO:8 #1.
XX
KW Tumour antigen; vaccine; immunoglobulin; CH3 fragment; fusion protein;
KW cancer; tumour; dendritic cell; endocytosis; immune response; cytostatic;
KW human; CAP-1; ds.
XX
OS Homo sapiens.
XX
PN WO2004024181-A1.
XX
PD 25-MAR-2004.
XX
PF 15-SEP-2003; 2003WO-CN000776.
XX
PR 13-SEP-2002; 2002CN-00136965.
XX
PA (LIJJ/) LI J.
XX
PI Li J;
XX
DR WPI; 2004-269898/25.
XX
PT Tumor-antigen vaccines with molecular weight far smaller than antigen-
PT antibody compound to enable endocytosis by dendritic cells to promote
PT very high immunoreaction for killing tumor cells.
XX
PS Example 2; SEQ ID NO 8; 28pp; Chinese.
XX
CC The invention relates to a tumour antigen vaccine comprising 7 or more
CC amino acids of a tumour antigen sequence joined to an immunoglobulin CH3
CC fragment. The invention also relates to DNA sequences encoding the
CC antigenic fusion polypeptide; expression vectors and host cells
CC comprising the DNA sequences; a process for recombinantly producing the
CC fusion polypeptide; and a vaccine composition comprising the fusion
CC polypeptide and a pharmaceutically acceptable carrier. The tumour antigen
CC used is preferably selected from 07-AP, AFP, ART-4, BAGE B, beta-
CC catenin/m, bcr-abl, CAMEL, CAP-1, CASP-8, CDC27m, CDK4/m, CEA, CT, Cyp-B,
CC DAM, ELF2M, ETV6-AML1, ETS, G250, GAGE, GnT-V, GP100, HAGE, HER- 2/NEU,
CC HLA-A*0201-R170I, HPV-E6, HPV-E7, EBNA, HSP70-2M, HST-2, hTERT, iCE,
CC KIAA0205, LAGE, LDLR/FUT, GDP-Lfucose, MAGE, MART-1/Melan-A, MCIR,
CC Myosin/m, MUC1, MUM-1,-2,-3, NA88-A, NY-ESO-1, P15, p190, P53, Pml/RAR
CC alpha, FRAME, PSA, PSM, RAGE, RAS, RU1, RU2, SAGE, SART-1, SART-3,
CC TEL/AML1, TPI/m, TRP-1, gp75, TRP-2, TRP-2/INT2 and WTI. The tumour
CC antigen vaccines of the invention are useful in cancer therapy. The
CC antigenic fusion protein used in the vaccine are much smaller than the
CC corresponding antibody-antigen complex, permitting them to be endocytosed
CC by dendritic cells and thereby resulting in a greatly increased anti-
CC tumour immune response. The present sequence represents DNA encoding a
CC fragment of the human CAP-1/CEA tumour antigen used in an example of the
CC invention. Note: The present sequence differs from that also referred to
CC as SEQ ID NO:8 () which is given on page 10 of the specification.
XX
SQ Sequence 30 BP; 6 A; 12 C; 5 G; 7 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	0.106	Length:	30
Score:	43.00	Matches:	8
Percent Similarity:	100.00%	Conservative:	1
Best Local Similarity:	88.89%	Mismatches:	0
Query Match:	95.56%	Indels:	0
DB:	12	Gaps:	0

US-10-725-373-4 (1-9) x ADL46174 (1-30)

```
Qy          1 TyrLeuSerGlyAlaAsnIleAsnLeu 9
              |||||||||||||||||:::||||||
Db          1 TACCTTTCGGGAGCGAACCTCAACCTC 27
```

RESULT 3
ADL46175
ID ADL46175 standard; DNA; 64 BP.
XX
AC ADL46175;
XX
DT 17-JUN-2004 (first entry)
XX
DE Human immunoglobulin Fc fragment 5' PCR primer, SEQ ID NO:9 #1.
XX
KW Tumour antigen; vaccine; immunoglobulin; CH3 fragment; fusion protein;
KW cancer; tumour; dendritic cell; endocytosis; immune response; cytostatic;
KW human; Fc fragment; PCR; primer; ss.
XX
OS Homo sapiens.
XX
PN WO2004024181-A1.
XX
PD 25-MAR-2004.
XX
PF 15-SEP-2003; 2003WO-CN000776.
XX
PR 13-SEP-2002; 2002CN-00136965.
XX
PA (LIJJ/) LI J.
XX
PI Li J;
XX
DR WPI; 2004-269898/25.
XX
PT Tumor-antigen vaccines with molecular weight far smaller than antigen-
PT antibody compound to enable endocytosis by dendritic cells to promote
PT very high immunoreaction for killing tumor cells.

XX
PS Example 2; SEQ ID NO 9; 28pp; Chinese.
XX
CC The invention relates to a tumour antigen vaccine comprising 7 or more
CC amino acids of a tumour antigen sequence joined to an immunoglobulin CH3
CC fragment. The invention also relates to DNA sequences encoding the
CC antigenic fusion polypeptide; expression vectors and host cells
CC comprising the DNA sequences; a process for recombinantly producing the
CC fusion polypeptide; and a vaccine composition comprising the fusion
CC polypeptide and a pharmaceutically acceptable carrier. The tumour antigen
CC used is preferably selected from 07-AP, AFP, ART-4, BAGE B, beta-
CC catenin/m, bcr-abl, CAMEL, CAP-1, CASP-8, CDC27m, CDK4/m, CEA, CT, Cyp-B,
CC DAM, ELF2M, ETV6-AML1, ETS, G250, GAGE, GnT-V, GP100, HAGE, HER- 2/NEU,
CC HLA-A*0201-R170I, HPV-E6, HPV-E7, EBNA, HSP70-2M, HST-2, hTERT, iCE,
CC KIAA0205, LAGE, LDLR/FUT, GDP-Lfucose, MAGE, MART-1/Melan-A, MCIR,
CC Myosin/m, MUC1, MUM-1,-2,-3, NA88-A, NY-ESO-1, P15, p190, P53, Pml/RAR
CC alpha, FRAME, PSA, PSM, RAGE, RAS, RU1, RU2, SAGE, SART-1, SART-3,
CC TEL/AML1, TPI/m, TRP-1, gp75, TRP-2, TRP-2/INT2 and WTI. The tumour
CC antigen vaccines of the invention are useful in cancer therapy. The
CC antigenic fusion protein used in the vaccine are much smaller than the
CC corresponding antibody-antigen complex, permitting them to be endocytosed
CC by dendritic cells and thereby resulting in a greatly increased anti-
CC tumour immune response. Sequences ADL46175-ADL46176 represent PCR primers
CC used to amplify DNA encoding a human immunoglobulin Fc fragment in an
CC example of the invention. Note: The present sequence differs from that
CC also referred to as SEQ ID NO:9 () which is given on page 10 of the
CC specification.
XX
SQ Sequence 64 BP; 16 A; 22 C; 13 G; 13 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	0.257	Length:	64
Score:	43.00	Matches:	8
Percent Similarity:	100.00%	Conservative:	1
Best Local Similarity:	88.89%	Mismatches:	0
Query Match:	95.56%	Indels:	0
DB:	12	Gaps:	0

US-10-725-373-4 (1-9) x ADL46175 (1-64)

```
Qy          1 TyrLeuSerGlyAlaAsnIleAsnLeu 9
              |||||||||||||||||:::||||||
Db         15 TACCTTTCGGGAGCGAACCTCAACCTC 41
```

RESULT 4
AAV57948/c
ID AAV57948 standard; DNA; 80 BP.
XX
AC AAV57948;
XX
DT 24-NOV-1998 (first entry)
XX
DE 708 vkcea primary reaction 1.2 oligonucleotide vkcea592r.
XX
KW Hepatitis B surface antigen; HBsAg; MHC class II-restricted peptide;
KW vaccination; vaccine; MHC class I molecule; immune response; cancer;
KW major histocompatibility complex molecule; pathogenic organism;
KW viral disease; autoimmune condition; allergy; PCR primer; ss.
XX
OS Synthetic.
XX
PN WO9833523-A1.
XX
PD 06-AUG-1998.
XX
PF 02-FEB-1998; 98WO-GB000325.
XX
PR 31-JAN-1997; 97GB-00001999.
PR 05-JUL-1997; 97GB-00014182.
PR 07-AUG-1997; 97GB-00016620.
PR 07-AUG-1997; 97GB-00016641.
PR 21-NOV-1997; 97GB-00024584.
XX
PA (BIOV-) BIOVATION LTD.
XX
PI Carr FJ, Carter G;
XX
DR WPI; 1998-437178/37.
XX
PT Immunogenic molecules - comprising nucleic acid and polypeptide portion,
PT from both of which peptide for presentation on major histocompatibility
PT complex molecules can be derived.
XX
PS Example 10; Page 60; 87pp; English.
XX
CC A molecule has been developed which comprises: (a) a nucleic acid portion
CC from which at least one peptide for presentation of MHC class I or class
CC II molecules, or both, may be derived, and (b) a polypeptide portion,
CC from which at least 1 peptide for presentation on MHC class I or class II
CC molecules, or both, may be derived. Also described in the present
CC invention is another molecule comprising: (a) a nucleic acid portion from
CC which at least 1 peptide for presentation on MHC class I or class II
CC molecules, or both, may be derived, and (b) a polypeptide portion
CC comprising a recognition domain capable of targeting the molecule to an
CC antigen presenting cell (APC), where the polypeptide portion does not
CC comprise a specific antigen binding site. The molecules can be used to
CC induce immune responses to treat or prevent, e.g. diseases caused by
CC pathogenic organisms, cancers, viral disease, e.g. HIV or hepatitis
CC infection, autoimmune conditions, e.g. Grave's disease, multiple
CC sclerosis, systemic lupus erythematosus, diabetes mellitus, Kawasaki's
CC disease, rheumatoid arthritis or allergies, e.g. atopic dermatitis,
CC allergic rhinitis, allergic conjunctivitis, atopic asthma or eczema. The
CC combination of DNA and polypeptide in the same molecule can give rise not
CC only to a combination of MHC class I- and MHC class II-mediated immune
CC responses but also to an enhancement of these responses compared to the
CC responses given by either DNA or polypeptide alone. The present sequence
CC represents an oligonucleotide used in an example from the present
CC invention
XX
SQ Sequence 80 BP; 16 A; 28 C; 19 G; 17 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	0.334	Length:	80
Score:	43.00	Matches:	8
Percent Similarity:	100.00%	Conservative:	1
Best Local Similarity:	88.89%	Mismatches:	0
Query Match:	95.56%	Indels:	0
DB:	2	Gaps:	0

US-10-725-373-4 (1-9) x AAV57948 (1-80)

```
Qy          1 TyrLeuSerGlyAlaAsnIleAsnLeu 9
              |||||||||||||||||:::||||||
Db         68 TACCTGTCCGGCGCCAACCTGAACCTG 42
```

RESULT 5
AAV81101/c
ID AAV81101 standard; DNA; 80 BP.
XX
AC AAV81101;
XX
DT 03-MAR-1999 (first entry)
XX
DE Vaccine 2 708 Vl constructing long oligo VKCEA592R.
XX
KW Non-immunogenic; epitope; T-cell; immunogenicity; immune system; SK;
KW immunoglobulin; therapeutic; streptokinase; vaccine; 708; ss.
XX
OS Synthetic.
XX
PN WO9852976-A1.
XX
PD 26-NOV-1998.
XX
PF 21-MAY-1998; 98WO-GB001473.
XX

PR 21-MAY-1997; 97GB-00010480.
PR 31-JUL-1997; 97GB-00016197.
PR 28-NOV-1997; 97GB-00025270.
PR 02-DEC-1997; 97US-0067235P.
PR 14-APR-1998; 98GB-00007751.
XX
PA (BIOV-) BIOVATION LTD.
XX
PI Carr FJ;
XX
DR WPI; 1999-045301/04.
XX
PT Reducing immunogenicity of proteins - by modifying the amino acid
PT sequence of the protein to eliminate potential epitopes for T-cells of a
PT given species.
XX
PS Example 4; Fig 20; 77pp; English.
XX
CC The invention relates to a method for the production of non-immunogenic
CC proteins. The method comprises determining at least part of the amino
CC acid sequence of the protein; (b) identifying in the amino acid sequence
CC one or more potential epitopes for T-cells (T-cell epitopes) of the given
CC species; and (c) modifying the amino acid sequence to eliminate at least
CC one of the T-cell epitopes identified in step (b) thereby to eliminate or
CC reduce the immunogenicity of the protein when exposed to the immune
CC system of the given species. A method of analysing a pre-existing protein
CC to predict the basis for immunogenic responses is also provided. The
CC methods can be used particularly for reducing the immunogenicity of
CC immunoglobulins or therapeutic proteins, e.g. Streptokinase (SK). The
CC products can be used for diagnosis and therapy. Sequences AAV81090-110
CC represent oligonucleotides used for the construction of vaccine 2 708 Vh
CC and Vl
XX
SQ Sequence 80 BP; 16 A; 28 C; 19 G; 17 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	0.334	Length:	80
Score:	43.00	Matches:	8
Percent Similarity:	100.00%	Conservative:	1
Best Local Similarity:	88.89%	Mismatches:	0
Query Match:	95.56%	Indels:	0
DB:	2	Gaps:	0

US-10-725-373-4 (1-9) x AAV81101 (1-80)

```
Qy          1 TyrLeuSerGlyAlaAsnIleAsnLeu 9
              |||||||||||||||||:::||||||
Db         68 TACCTGTCCGGCGCCAACCTGAACCTG 42
```

RESULT 6
AAI29234
ID AAI29234 standard; cDNA; 155 BP.
XX
AC AAI29234;
XX
DT 12-OCT-2001 (first entry)
XX
DE Colon tumour related determined cDNA sequence for clone R0094:D08.
XX
KW Human; immunotherapy; diagnosis; colon cancer; colon tumour; immunogenic;
KW gene therapy; vaccine; colonic cancer; ss.
XX
OS Homo sapiens.
XX
PN WO200149716-A2.
XX
PD 12-JUL-2001.
XX
PF 29-DEC-2000; 2000WO-US035596.
XX
PR 30-DEC-1999; 99US-00476296.
PR 10-JAN-2000; 2000US-00480321.
PR 15-FEB-2000; 2000US-00504629.
PR 06-MAR-2000; 2000US-00519444.
PR 19-MAY-2000; 2000US-00575251.
PR 29-JUN-2000; 2000US-00609448.
PR 28-AUG-2000; 2000US-00649811.
XX
PA (CORI-) CORIXA CORP.
XX
PI Xu J, Lodes MJ, Secrist H, Benson DR, Meagher MJ, Stolk JA;
PI King GE, Wang T, Jiang Y;
XX
DR WPI; 2001-441847/47.
XX
PT Colon tumor associated proteins and nucleic acids useful for the
PT prevention, diagnosis and treatment of colonic cancer.
XX
PS Claim 2; Page 356; 472pp; English.
XX
CC The present invention describes colon tumour associated proteins (I) and
CC the polynucleotides (II) that encode them. (I) have cytostatic activity.
CC (I) and (II) can be used in gene therapy and vaccine production. (I) and
CC (II) may be used in the prevention, diagnosis and treatment of diseases
CC associated with inappropriate colon tumour associated protein (TCAP)
CC expression, such as colonic cancer. For example, (I) and (II) may be used
CC to treat disorders associated with decreased expression by rectifying
CC mutations or deletions in a patient's genome that affect the activity of
CC TCAPs by expressing inactive proteins or to supplement the patients own
CC production of them. Additionally, (II) may be used to produce the TCAP
CC proteins, by inserting the nucleic acids into a host cell culturing the
CC cell to express the protein. (II) and its complementary sequences may
CC also be used as DNA probes in diagnostic polymerase chain reaction (PCR)
CC and hybridisation assays to detect and quantitate the presence of similar
CC nucleic acids in samples, and therefore which patients may be in need of
CC restorative therapy. (I) may also be used as antigens in the production
CC of antibodies against TCAPs and in assays to identify modulators of TCAP
CC expression and activity. Anti-(I) antibodies and antagonists may also be
CC used to down regulate TCAP expression and activity. The anti-(I)
CC antibodies may also be used as diagnostic agents for detecting the
CC presence of TCAPs in samples (e.g. by enzyme linked immunosorbant assay
CC (ELISA)). AAI28460 to AAI29512 and AAM24494 to AAM24523 represent
CC nucleotide and amino acid sequences given in the exemplification of the
CC present invention
XX
SQ Sequence 155 BP; 30 A; 56 C; 31 G; 30 T; 0 U; 8 Other;

Alignment Scores:

Pred. No.:	0.727	Length:	155
Score:	43.00	Matches:	8
Percent Similarity:	100.00%	Conservative:	1
Best Local Similarity:	88.89%	Mismatches:	0
Query Match:	95.56%	Indels:	0
DB:	4	Gaps:	0

US-10-725-373-4 (1-9) x AAI29234 (1-155)

```
Qy          1 TyrLeuSerGlyAlaAsnIleAsnLeu 9
              |||||||||||||||||:::||||||
Db        119 TACCTTTCNGGAGCGAACCTCAACCTC 145
```

RESULT 7
ABZ33420
ID ABZ33420 standard; cDNA; 155 BP.
XX
AC ABZ33420;
XX
DT 30-JAN-2003 (first entry)
XX
DE Human colon tumour cDNA for clone R0094:D08 SEQ ID NO:788.
XX
KW Human; colon cancer; colon tumour; immunotherapy; diagnosis; cancer;
KW tumour; immune response; immunostimulant; cytostatic; vaccine; gene; ss.
XX
OS Homo sapiens.
XX

PN WO200283070-A2.
XX
PD 24-OCT-2002.
XX
PF 09-APR-2002; 2002WO-US011475.
XX
PR 10-APR-2001; 2001US-00833263.
PR 03-AUG-2001; 2001US-00922217.
PR 19-DEC-2001; 2001US-00025380.
XX
PA (CORI-) CORIXA CORP.
XX
PI Xu J, Lodes MJ, Secrist H, Benson DR, Meagher MJ, Stolk JA;
PI Wang T, Jiang Y, Smith CL, King GE, Wang A, Clapper JD, Skeiky YAW;
PI Fanger GR, Vedvick TS, Carter D;
XX
DR WPI; 2003-067548/06.
XX
PT New polynucleotide, useful for the preparation of a composition for
PT stimulating an immune response against, or treating, cancer.
XX
PS Disclosure; Page 357; 537pp; English.
XX
CC The present invention describes compounds (I) for the immunotherapy and
CC diagnosis of colon cancer. Also described: (1) a method for detecting the
CC presence of cancer in a patient; (2) a method for stimulating and/or
CC expanding T cells specific for a tumour protein; (3) an isolated T cell
CC population comprising T cells prepared by the method of (2); (4) a method
CC for stimulating an immune response in a patient; (5) a method for
CC treating cancer in a patient; and (6) a method for inhibiting the
CC development of cancer in a patient. (I) have immunostimulant and
CC cytostatic activities and can be used in vaccines. ABZ32646 to ABZ33725
CC and ABP55343 to ABP55391 represent human colon cancer/tumour related
CC sequences used in the exemplification of the present invention
XX
SQ Sequence 155 BP; 30 A; 56 C; 31 G; 30 T; 0 U; 8 Other;

Alignment Scores:

Pred. No.:	0.727	Length:	155
Score:	43.00	Matches:	8
Percent Similarity:	100.00%	Conservative:	1
Best Local Similarity:	88.89%	Mismatches:	0
Query Match:	95.56%	Indels:	0
DB:	8	Gaps:	0

US-10-725-373-4 (1-9) x ABZ33420 (1-155)

```
Qy          1 TyrLeuSerGlyAlaAsnIleAsnLeu 9
              |||||||||||||||||:::||||||
Db        119 TACCTTTCNGGAGCGAACCTCAACCTC 145
```

RESULT 8
AAS57750
ID AAS57750 standard; cDNA; 256 BP.
XX
AC AAS57750;
XX
DT 13-FEB-2002 (first entry)
XX
DE cDNA #426 encoding portion of a human colon tumour protein.
XX
KW Human; colon tumour protein; colon cancer; gene therapy; cytostatic; ss.
XX
OS Homo sapiens.
XX
PN WO200173027-A2.
XX
PD 04-OCT-2001.
XX
PF 22-MAR-2001; 2001WO-US009246.
XX
PR 24-MAR-2000; 2000US-0191597P.
PR 04-MAY-2000; 2000US-0202024P.
PR 05-MAY-2000; 2000US-0202189P.
XX
PA (CORI-) CORIXA CORP.
XX
PI Meagher MJ, Xu J, King GE;
XX
DR WPI; 2001-611627/70.
XX
PT New colon tumor proteins and related nucleic acid, useful for treatment,
PT prevention, diagnosis and monitoring of cancer.
XX
PS Claim 4; Page 125; 299pp; English.
XX
CC Th present invention relates to the isolation of novel cDNA sequences
CC encoding for at least an immunogenic portion of human colon tumour
CC proteins. The sequences of the invention are useful in pharmaceutical
CC compositions and vaccines for the prevention and treatment of cancers
CC such as colon cancer. They are also useful for the diagnosis and
CC monitoring of such cancers. Antibodies to the colon tumour proteins and
CC antigen presenting cells that express polynucleotides encoding colon
CC tumour proteins can be used to inhibit the development of cancers. T-
CC cells that react specifically with colon tumour proteins are useful for
CC removing tumour cells from samples (e.g. blood) and for cancer treatment.
CC The polynucleotides sequences are also useful in gene therapy. AAS57325-
CC AAS58880 represent the cDNA sequences of the invention that encode for
CC portions of human colon tumour proteins
XX
SQ Sequence 256 BP; 62 A; 84 C; 44 G; 66 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	1.31	Length:	256
Score:	43.00	Matches:	8
Percent Similarity:	100.00%	Conservative:	1
Best Local Similarity:	88.89%	Mismatches:	0
Query Match:	95.56%	Indels:	0
DB:	4	Gaps:	0

US-10-725-373-4 (1-9) x AAS57750 (1-256)

```
Qy          1 TyrLeuSerGlyAlaAsnIleAsnLeu 9
              |||||||||||||||||:::||||||
Db         15 TACCTTTCGGGAGCGAACCTCAACCTC 41
```

RESULT 9
ABV88334/c
ID ABV88334 standard; cDNA; 340 BP.
XX
AC ABV88334;
XX
DT 13-DEC-2002 (first entry)
XX
DE Human colon cancer related cDNA SEQ ID NO 1645.
XX
KW Human; colon; cancer; cytostatic; tumour; gene therapy; vaccine; gene;
KW ss.
XX
OS Homo sapiens.
XX
PN WO200258534-A2.
XX
PD 01-AUG-2002.
XX
PF 16-NOV-2001; 2001WO-US043704.
XX
PR 20-NOV-2000; 2000US-0252222P.
PR 06-FEB-2001; 2001US-0267011P.
PR 28-MAR-2001; 2001US-0279670P.
PR 10-JUL-2001; 2001US-0304037P.
XX
PA (CORI-) CORIXA CORP.
XX
PI Stolk JA, Xu J, Chenault RA, Meagher MJ, Secrist H, King GE;
XX

```
DR   WPI; 2002-608400/65.
XX
PT   New isolated tumor colon polynucleotide and polypeptide, useful for the
PT   diagnosis, prevention and/or treatment of cancer, in particular colon
PT   cancer.
XX
PS   Claim 1; SEQ ID NO 1645; 266pp + Sequence Listing; English.
XX
CC   The invention relates to a human colon tumour expressed polynucleotide
CC   (I) encoding a polypeptide (II, ABP67991-ABP67996) comprising: (i) any of
CC   2600 fully defined nucleotide sequences (ABV8669-ABV89289); (ii)
CC   complements of (i); (iii) at least 20 contiguous residues of (i); (iv)
CC   sequences that hybridize to (i), under moderately stringent conditions;
CC   (v) sequences having at least 75% or 90% identity to (i); or (vi)
CC   degenerate variants of (i). The compositions and methods of the present
CC   invention are useful for the diagnosis, prevention and/or treatment of
CC   cancer, particularly colon cancer. (I) can be used in gene therapy and
CC   (I) and (II) are useful in pharmaceutical compositions such as vaccines.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 340 BP; 82 A; 63 C; 114 G; 81 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:              1.83          Length:         340
Score:                  43.00         Matches:        8
Percent Similarity:     100.00%       Conservative:   1
Best Local Similarity:  88.89%        Mismatches:     0
Query Match:            95.56%        Indels:         0
DB:                     6             Gaps:           0

US-10-725-373-4 (1-9) x ABV88334 (1-340)

Qy          1 TyrLeuSerGlyAlaAsnIleAsnLeu 9
              |||||||||||||||||||:::||||||
Db        241 TACCTTTCGGGAGCGAACCTCAACCTC 215

RESULT 10
AAS57366/c
ID   AAS57366 standard; cDNA; 407 BP.
XX
AC   AAS57366;
XX
DT   13-FEB-2002  (first entry)
XX
DE   cDNA #42 encoding portion of a human colon tumour protein.
XX
KW   Human; colon tumour protein; colon cancer; gene therapy; cytostatic; ss.
XX
OS   Homo sapiens.
XX
PN   WO200173027-A2.
XX
PD   04-OCT-2001.
XX
PF   22-MAR-2001; 2001WO-US009246.
XX
PR   24-MAR-2000; 2000US-0191597P.
PR   04-MAY-2000; 2000US-0202024P.
PR   05-MAY-2000; 2000US-0202189P.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Meagher MJ,  Xu J,  King GE;
XX
DR   WPI; 2001-611627/70.
XX
PT   New colon tumor proteins and related nucleic acid, useful for treatment,
PT   prevention, diagnosis and monitoring of cancer.
XX
PS   Claim 4; Page 68; 299pp; English.
XX
CC   Th present invention relates to the isolation of novel cDNA sequences
CC   encoding for at least an immunogenic portion of human colon tumour
CC   proteins. The sequences of the invention are useful in pharmaceutical
CC   compositions and vaccines for the prevention and treatment of cancers
CC   such as colon cancer. They are also useful for the diagnosis and
CC   monitoring of such cancers. Antibodies to the colon tumour proteins and
CC   antigen presenting cells that express polynucleotides encoding colon
CC   tumour proteins can be used to inhibit the development of cancers. T-
CC   cells that react specifically with colon tumour proteins are useful for
CC   removing tumour cells from samples (e.g. blood) and for cancer treatment.
CC   The polynucleotides sequences are also useful in gene therapy. AAS57325-
CC   AAS58880 represent the cDNA sequences of the invention that encode for
CC   portions of human colon tumour proteins
XX
SQ   Sequence 407 BP; 97 A; 76 C; 130 G; 100 T; 0 U; 4 Other;

Alignment Scores:
Pred. No.:              2.26          Length:         407
Score:                  43.00         Matches:        8
Percent Similarity:     100.00%       Conservative:   1
Best Local Similarity:  88.89%        Mismatches:     0
Query Match:            95.56%        Indels:         0
DB:                     4             Gaps:           0

US-10-725-373-4 (1-9) x AAS57366 (1-407)

Qy          1 TyrLeuSerGlyAlaAsnIleAsnLeu 9
              |||||||||||||||||||:::||||||
Db        241 TACCTTTCGGGAGCGAACCTCAACCTC 215

RESULT 11
AAS57425/c
ID   AAS57425 standard; cDNA; 409 BP.
XX
AC   AAS57425;
XX
DT   13-FEB-2002  (first entry)
XX
DE   cDNA #101 encoding portion of a human colon tumour protein.
XX
KW   Human; colon tumour protein; colon cancer; gene therapy; cytostatic; ss.
XX
OS   Homo sapiens.
XX
PN   WO200173027-A2.
XX
PD   04-OCT-2001.
XX
PF   22-MAR-2001; 2001WO-US009246.
XX
PR   24-MAR-2000; 2000US-0191597P.
PR   04-MAY-2000; 2000US-0202024P.
PR   05-MAY-2000; 2000US-0202189P.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Meagher MJ,  Xu J,  King GE;
XX
DR   WPI; 2001-611627/70.
XX
PT   New colon tumor proteins and related nucleic acid, useful for treatment,
PT   prevention, diagnosis and monitoring of cancer.
XX
PS   Claim 4; Page 77; 299pp; English.
XX
CC   Th present invention relates to the isolation of novel cDNA sequences
CC   encoding for at least an immunogenic portion of human colon tumour
CC   proteins. The sequences of the invention are useful in pharmaceutical
CC   compositions and vaccines for the prevention and treatment of cancers
CC   such as colon cancer. They are also useful for the diagnosis and
CC   monitoring of such cancers. Antibodies to the colon tumour proteins and
CC   antigen presenting cells that express polynucleotides encoding colon
CC   tumour proteins can be used to inhibit the development of cancers. T-
```

```
CC   cells that react specifically with colon tumour proteins are useful for
CC   removing tumour cells from samples (e.g. blood) and for cancer treatment.
CC   The polynucleotides sequences are also useful in gene therapy. AAS57325-
CC   AAS58880 represent the cDNA sequences of the invention that encode for
CC   portions of human colon tumour proteins
XX
SQ   Sequence 409 BP; 98 A; 76 C; 132 G; 103 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:              2.27          Length:         409
Score:                  43.00         Matches:        8
Percent Similarity:     100.00%       Conservative:   1
Best Local Similarity:  88.89%        Mismatches:     0
Query Match:            95.56%        Indels:         0
DB:                     4             Gaps:           0

US-10-725-373-4 (1-9) x AAS57425 (1-409)

Qy          1 TyrLeuSerGlyAlaAsnIleAsnLeu 9
              |||||||||||||||||:::||||||
Db        241 TACCTTTCGGGAGCGAACCTCAACCTC 215

RESULT 12
ABV86774
ID   ABV86774 standard; cDNA; 409 BP.
XX
AC   ABV86774;
XX
DT   13-DEC-2002  (first entry)
XX
DE   Human colon cancer related cDNA SEQ ID NO 85.
XX
KW   Human; colon; cancer; cytostatic; tumour; gene therapy; vaccine; gene;
KW   ss.
XX
OS   Homo sapiens.
XX
PN   WO200258534-A2.
XX
PD   01-AUG-2002.
XX
PF   16-NOV-2001; 2001WO-US043704.
XX
PR   20-NOV-2000; 2000US-0252222P.
PR   06-FEB-2001; 2001US-0267011P.
PR   28-MAR-2001; 2001US-0279670P.
PR   10-JUL-2001; 2001US-0304037P.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Stolk JA,  Xu J,  Chenault RA,  Meagher MJ,  Secrist H,  King GE;
XX
DR   WPI; 2002-608400/65.
XX
PT   New isolated tumor colon polynucleotide and polypeptide, useful for the
PT   diagnosis, prevention and/or treatment of cancer, in particular colon
PT   cancer.
XX
PS   Claim 1; SEQ ID NO 85; 266pp + Sequence Listing; English.
XX
CC   The invention relates to a human colon tumour expressed polynucleotide
CC   (I) encoding a polypeptide (II, ABP67991-ABP67996) comprising: (i) any of
CC   2600 fully defined nucleotide sequences (ABV8669-ABV89289); (ii)
CC   complements of (i); (iii) at least 20 contiguous residues of (i); (iv)
CC   sequences that hybridize to (i), under moderately stringent conditions;
CC   (v) sequences having at least 75% or 90% identity to (i); or (vi)
CC   degenerate variants of (i). The compositions and methods of the present
CC   invention are useful for the diagnosis, prevention and/or treatment of
CC   cancer, particularly colon cancer. (I) can be used in gene therapy and
CC   (I) and (II) are useful in pharmaceutical compositions such as vaccines.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 409 BP; 103 A; 132 C; 76 G; 98 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:              2.27          Length:         409
Score:                  43.00         Matches:        8
Percent Similarity:     100.00%       Conservative:   1
Best Local Similarity:  88.89%        Mismatches:     0
Query Match:            95.56%        Indels:         0
DB:                     6             Gaps:           0

US-10-725-373-4 (1-9) x ABV86774 (1-409)

Qy          1 TyrLeuSerGlyAlaAsnIleAsnLeu 9
              |||||||||||||||||:::||||||
Db        169 TACCTTTCGGGAGCGAACCTCAACCTC 195

RESULT 13
ABV87551/c
ID   ABV87551 standard; cDNA; 409 BP.
XX
AC   ABV87551;
XX
DT   13-DEC-2002  (first entry)
XX
DE   Human colon cancer related cDNA SEQ ID NO 862.
XX
KW   Human; colon; cancer; cytostatic; tumour; gene therapy; vaccine; gene;
KW   ss.
XX
OS   Homo sapiens.
XX
PN   WO200258534-A2.
XX
PD   01-AUG-2002.
XX
PF   16-NOV-2001; 2001WO-US043704.
XX
PR   20-NOV-2000; 2000US-0252222P.
PR   06-FEB-2001; 2001US-0267011P.
PR   28-MAR-2001; 2001US-0279670P.
PR   10-JUL-2001; 2001US-0304037P.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Stolk JA,  Xu J,  Chenault RA,  Meagher MJ,  Secrist H,  King GE;
XX
DR   WPI; 2002-608400/65.
XX
PT   New isolated tumor colon polynucleotide and polypeptide, useful for the
PT   diagnosis, prevention and/or treatment of cancer, in particular colon
PT   cancer.
XX
PS   Claim 1; SEQ ID NO 862; 266pp + Sequence Listing; English.
XX
CC   The invention relates to a human colon tumour expressed polynucleotide
CC   (I) encoding a polypeptide (II, ABP67991-ABP67996) comprising: (i) any of
CC   2600 fully defined nucleotide sequences (ABV8669-ABV89289); (ii)
CC   complements of (i); (iii) at least 20 contiguous residues of (i); (iv)
CC   sequences that hybridize to (i), under moderately stringent conditions;
CC   (v) sequences having at least 75% or 90% identity to (i); or (vi)
CC   degenerate variants of (i). The compositions and methods of the present
CC   invention are useful for the diagnosis, prevention and/or treatment of
CC   cancer, particularly colon cancer. (I) can be used in gene therapy and
CC   (I) and (II) are useful in pharmaceutical compositions such as vaccines.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 409 BP; 98 A; 76 C; 132 G; 103 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:              2.27          Length:         409
```

```
Score:                  43.00          Matches:       8
Percent Similarity:     100.00%        Conservative:  1
Best Local Similarity:  88.89%         Mismatches:    0
Query Match:            95.56%         Indels:        0
DB:                     6              Gaps:          0

US-10-725-373-4 (1-9) x ABV87551 (1-409)

Qy          1 TyrLeuSerGlyAlaAsnIleAsnLeu 9
              |||||||||||||||||||:::||||||
Db        241 TACCTTTCGGGAGCGAACCTCAACCTC 215

RESULT 14
ABV89100
ID   ABV89100 standard; cDNA; 409 BP.
XX
AC   ABV89100;
XX
DT   13-DEC-2002  (first entry)
XX
DE   Human colon cancer related cDNA SEQ ID NO 2411.
XX
KW   Human; colon; cancer; cytostatic; tumour; gene therapy; vaccine; gene;
KW   ss.
XX
OS   Homo sapiens.
XX
PN   WO200258534-A2.
XX
PD   01-AUG-2002.
XX
PF   16-NOV-2001; 2001WO-US043704.
XX
PR   20-NOV-2000; 2000US-0252222P.
PR   06-FEB-2001; 2001US-0267011P.
PR   28-MAR-2001; 2001US-0279670P.
PR   10-JUL-2001; 2001US-0304037P.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Stolk JA,  Xu J,  Chenault RA,  Meagher MJ,  Secrist H,  King GE;
XX
DR   WPI; 2002-608400/65.
XX
PT   New isolated tumor colon polynucleotide and polypeptide, useful for the
PT   diagnosis, prevention and/or treatment of cancer, in particular colon
PT   cancer.
XX
PS   Claim 1; SEQ ID NO 2411; 266pp + Sequence Listing; English.
XX
CC   The invention relates to a human colon tumour expressed polynucleotide
CC   (I) encoding a polypeptide (II, ABP67991-ABP67996) comprising: (i) any of
CC   2600 fully defined nucleotide sequences (ABV8669-ABV89289); (ii)
CC   complements of (i); (iii) at least 20 contiguous residues of (i); (iv)
CC   sequences that hybridize to (i), under moderately stringent conditions;
CC   (v) sequences having at least 75% or 90% identity to (i); or (vi)
CC   degenerate variants of (i). The compositions and methods of the present
CC   invention are useful for the diagnosis, prevention and/or treatment of
CC   cancer, particularly colon cancer. (I) can be used in gene therapy and
CC   (I) and (II) are useful in pharmaceutical compositions such as vaccines.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 409 BP; 103 A; 131 C; 76 G; 99 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:              2.27           Length:        409
Score:                  43.00          Matches:       8
Percent Similarity:     100.00%        Conservative:  1
Best Local Similarity:  88.89%         Mismatches:    0
Query Match:            95.56%         Indels:        0
DB:                     6              Gaps:          0

US-10-725-373-4 (1-9) x ABV89100 (1-409)

Qy          1 TyrLeuSerGlyAlaAsnIleAsnLeu 9
              |||||||||||||||||||:::||||||
Db        169 TACCTTTCGGGAGCGAACCTCAACCTC 195

RESULT 15
ABV87855
ID   ABV87855 standard; cDNA; 409 BP.
XX
AC   ABV87855;
XX
DT   13-DEC-2002  (first entry)
XX
DE   Human colon cancer related cDNA SEQ ID NO 1166.
XX
KW   Human; colon; cancer; cytostatic; tumour; gene therapy; vaccine; gene;
KW   ss.
XX
OS   Homo sapiens.
XX
PN   WO200258534-A2.
XX
PD   01-AUG-2002.
XX
PF   16-NOV-2001; 2001WO-US043704.
XX
PR   20-NOV-2000; 2000US-0252222P.
PR   06-FEB-2001; 2001US-0267011P.
PR   28-MAR-2001; 2001US-0279670P.
PR   10-JUL-2001; 2001US-0304037P.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Stolk JA,  Xu J,  Chenault RA,  Meagher MJ,  Secrist H,  King GE;
XX
DR   WPI; 2002-608400/65.
XX
PT   New isolated tumor colon polynucleotide and polypeptide, useful for the
PT   diagnosis, prevention and/or treatment of cancer, in particular colon
PT   cancer.
XX
PS   Claim 1; SEQ ID NO 1166; 266pp + Sequence Listing; English.
XX
CC   The invention relates to a human colon tumour expressed polynucleotide
CC   (I) encoding a polypeptide (II, ABP67991-ABP67996) comprising: (i) any of
CC   2600 fully defined nucleotide sequences (ABV8669-ABV89289); (ii)
CC   complements of (i); (iii) at least 20 contiguous residues of (i); (iv)
CC   sequences that hybridize to (i), under moderately stringent conditions;
CC   (v) sequences having at least 75% or 90% identity to (i); or (vi)
CC   degenerate variants of (i). The compositions and methods of the present
CC   invention are useful for the diagnosis, prevention and/or treatment of
CC   cancer, particularly colon cancer. (I) can be used in gene therapy and
CC   (I) and (II) are useful in pharmaceutical compositions such as vaccines.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 409 BP; 103 A; 132 C; 76 G; 98 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:              2.27           Length:        409
Score:                  43.00          Matches:       8
Percent Similarity:     100.00%        Conservative:  1
Best Local Similarity:  88.89%         Mismatches:    0
Query Match:            95.56%         Indels:        0
DB:                     6              Gaps:          0

US-10-725-373-4 (1-9) x ABV87855 (1-409)

Qy          1 TyrLeuSerGlyAlaAsnIleAsnLeu 9
              |||||||||||||||||||:::||||||
```

Db 169 TACCTTTCGGGAGCGAACCTCAACCTC 195

Search completed: May 17, 2005, 17:45:08
Job time : 326.5 secs

OM protein - protein search, using sw model

Run on: May 17, 2005, 06:13:14 ; Search time 11.25 Seconds
(without alignments)
76.973 Million cell updates/sec

Title: US-10-725-373-5
Perfect score: 48
Sequence: 1 YLSGACLNL 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_79:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	36	75.0	702	2	A36319	carcinoembryonic a
2	36	75.0	3328	2	T30835	breast cancer tumo
3	36	75.0	3329	2	T42205	breast cancer susc
4	36	75.0	3329	2	T30904	breast cancer tumo
5	35	72.9	57	2	S19088	dihydrolipoamide d
6	35	72.9	314	2	S56055	hypothetical prote
7	35	72.9	4302	2	A38971	polycystic kidney
8	34	70.8	91	2	I68703	MHC c6/g2 (Qa-2) p
9	34	70.8	163	2	H84039	hypothetical prote
10	34	70.8	219	2	B49181	beta B2-crystallin
11	34	70.8	254	2	H89772	hypothetical prote
12	34	70.8	265	2	B84468	hypothetical prote
13	34	70.8	451	2	S15236	glutathione-disulf
14	34	70.8	2395	1	S50820	surface protein ty
15	33	68.8	114	2	T29869	hypothetical prote
16	33	68.8	261	2	JC5806	aquaporin 8 - mous
17	33	68.8	263	2	JC5622	aquaporin 8 - rat
18	33	68.8	345	2	S07114	MHC class I histoc
19	33	68.8	348	2	T18230	alcohol dehydrogen
20	33	68.8	348	2	S29990	histocompatibility
21	33	68.8	362	2	A60384	MHC class I histoc
22	33	68.8	363	2	T42371	probable mannose-1
23	33	68.8	363	2	T41209	mannose-1-phosphat
24	33	68.8	365	2	H90113	nucleolar snRNP pr
25	33	68.8	368	2	E64142	hypothetical prote
26	33	68.8	453	2	C86176	hypothetical prote
27	33	68.8	457	2	I64229	dihydrolipoamide d
28	33	68.8	457	2	S73774	dihydrolipoamide d
29	33	68.8	637	1	YCRP	acetolactate synth
30	33	68.8	659	2	S60056	acetolactate synth
31	33	68.8	659	2	S60058	acetolactate synth
32	33	68.8	670	1	YCMU	acetolactate synth
33	33	68.8	933	2	E64603	hypothetical prote
34	33	68.8	1017	2	T30542	major surface glyc
35	33	68.8	2210	1	RRXPLC	genome polyprotein
36	32	66.7	140	2	H36851	RNA-binding ribonu
37	32	66.7	140	2	T28576	6R protein - vario
38	32	66.7	146	2	B81260	ribonuclease H (EC
39	32	66.7	146	2	H72167	A35R protein - var
40	32	66.7	249	1	CYRTB1	beta-crystallin B1
41	32	66.7	253	2	S07264	beta-crystallin B1
42	32	66.7	276	2	D70081	conserved hypothet
43	32	66.7	277	2	S26008	hypothetical prote
44	32	66.7	286	2	F86293	T24D18.2 protein -
45	32	66.7	288	2	S44256	fructokinase (EC 2

ALIGNMENTS

RESULT 1
A36319
carcinoembryonic antigen precursor - human
N;Alternate names: CEA; meconium antigen 100
C;Species: Homo sapiens (man)
C;Date: 16-Sep-1992 #sequence_revision 16-Sep-1992 #text_change 09-Jul-2004
C;Accession: A36319; A27773; A31037; A25845; S08106; S31737; A44476; I54224; I59098; A26
R;Schrewe, H.; Thompson, J.; Bona, M.; Hefta, L.J.F.; Maruya, A.; Hassauer, M.; Shively,
Mol. Cell. Biol. 10, 2738-2748, 1990
A;Title: Cloning of the complete gene for carcinoembryonic antigen: analysis of its promc
A;Reference number: A36319; MUID:90258861; PMID:2342461
A;Accession: A36319
A;Molecule type: DNA
A;Residues: 1-702 <SCH>
A;Cross-references: UNIPROT:P06731; GB:M17303; NID:g178676; PIDN:AAB59513.1; PID:g178677
A;Note: the authors show the codons TTA for residue 641-Phe and CAG for residue 646-Thr
R;Beauchemin, N.; Benchimol, S.; Cournoyer, D.; Fuks, A.; Stanners, C.P.
Mol. Cell. Biol. 7, 3221-3230, 1987
A;Title: Isolation and characterization of full-length functional cDNA clones for human c
A;Reference number: A27773; MUID:88038876; PMID:3670312
A;Accession: A27773
A;Molecule type: mRNA
A;Residues: 1-702 <BEA>
A;Cross-references: GB:M29540; NID:g180222; PIDN:AAA51967.1; PID:g180223
R;Barnett, T.; Goebel, S.J.; Nothdurft, M.A.; Elting, J.J.
Genomics 3, 59-66, 1988
A;Title: Carcinoembryonic antigen family: characterization of cDNAs coding for NCA and CE
A;Reference number: A31037; MUID:89122014; PMID:3220478
A;Accession: A31037
A;Molecule type: mRNA
A;Residues: 1-702 <BAR>
A;Cross-references: GB:M29540; NID:g180222; PIDN:AAA51967.1; PID:g180223
A;Note: the authors translated the codon GTG for residue 130 as Leu
R;Oikawa, S.; Nakazato, H.; Kosaki, G.
Biochem. Biophys. Res. Commun. 142, 511-518, 1987
A;Title: Primary structure of human carcinoembryonic antigen (CEA) deduced from cDNA sequ
A;Reference number: A25845; MUID:87128144; PMID:3814146
A;Accession: A25845
A;Molecule type: mRNA
A;Residues: 5-702 <OIK>
A;Cross-references: GB:M15042; NID:g180198; PIDN:AAA51963.1; PID:g180199
R;Oikawa, S.
submitted to the EMBL Data Library, September 1989
A;Reference number: S08106
A;Accession: S08106
A;Molecule type: mRNA
A;Residues: 5-319,321-702 <OI2>
A;Cross-references: EMBL:X16455; NID:g29854; PIDN:CAA34474.1; PID:g825638
R;Barnett, T.
submitted to the EMBL Data Library, September 1991
A;Description: Genomic DNA sequence upstream of the translational start of the carcinoemb
A;Reference number: S31737

A;Accession: S31737
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-141 <BA2>
A;Cross-references: EMBL:X62151
R;Khan, W.N.; Fraengsmyr, L.; Teglund, S.; Israelsson, A.; Bremer, K.; Hammarstroem, S.
Genomics 14, 384-390, 1992
A;Title: Identification of three new genes and estimation of the size of the carcinoembr
A;Reference number: A44476; MUID:93052339; PMID:1427854
A;Accession: A44476
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: DNA
A;Residues: 35-141 <KHA>
R;Willcocks, T.C.; Craig, I.W.
Genomics 8, 492-500, 1990
A;Title: Characterization of the genomic organization of human carcinoembryonic antigen
A;Reference number: I54224; MUID:91139118; PMID:2286372
A;Accession: I54224
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-37 <RES>
A;Cross-references: GB:M60964; NID:g180215; PIDN:AAA51964.1; PID:g180217
R;Zimmermann, W.; Ortlieb, B.; Friedrich, R.; von Kleist, S.
Proc. Natl. Acad. Sci. U.S.A. 84, 2960-2964, 1987
A;Title: Isolation and characterization of cDNA clones encoding the human carcinoembryon
A;Reference number: I59098; MUID:87204247; PMID:3033671
A;Accession: I59098
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 331-702 <RE2>
A;Cross-references: GB:M16234; NID:g180240; PIDN:AAA51972.1; PID:g180241
R;Siepen, D.; Paxton, R.J.; Neumaier, M.; Shively, J.E.; Wagener, C.
Biochem. Biophys. Res. Commun. 147, 212-218, 1987
A;Title: Carcinoembryonic antigen (CEA) and two crossreacting antigens of 165 KD and 105
A;Reference number: A26831; MUID:87326349; PMID:3632664
A;Accession: A26831
A;Molecule type: protein
A;Residues: 35-64 <SIE>
R;Thomas, P.; Toth, C.A.
Biochem. Biophys. Res. Commun. 170, 391-396, 1990
A;Title: Carcinoembryonic antigen binding to Kupffer cells is via a peptide located at t
A;Reference number: A35490; MUID:90321257; PMID:2372297
A;Accession: A35490
A;Molecule type: protein
A;Residues: 'X',140-151,'X',153,'X',155-156 <THO>
A;Note: this is the amino terminal end of a fragment shown to mediate uptake by Kupffer
C;Comment: This heavily glycosylated membane protein of unknown function is a widely use
C;Comment: This protein may be processed at its C-terminus. It is anchored to the membra
C;Genetics:
A;Gene: GDB:CEA
A;Cross-references: GDB:119054; OMIM:114890
A;Map position: 19q13.2-19q13.2
A;Introns: 22/1; 142/1; 235/1; 320/1; 413/1; 498/1; 591/1; 676/1
C;Superfamily: carcinoembryonic antigen; carcinoembryonic antigen precursor amino-termin
C;Keywords: blocked carboxyl end; glycoprotein; lipoprotein; membrane protein; phosphati
F;1-138/Domain: carcinoembryonic antigen precursor amino-terminal homology <CEAN>
F;1-34/Domain: signal sequence #status predicted <SIG>
F;35-678/Product: carcinoembryonic antigen #status predicted <MAT>
F;160-217/Domain: immunoglobulin homology <IMM1>
F;252-301/Domain: immunoglobulin homology <IMM2>
F;338-395/Domain: immunoglobulin homology <IMM3>
F;516-573/Domain: immunoglobulin homology <IMM4>
F;608-657/Domain: immunoglobulin homology <IMM5>
F;679-702/Domain: carboxyl-terminal propeptide #status predicted <CTP>
F;678/Modified site: GPI-anchor ethanolamine amidated carboxyl end (Gly) (in mature form

  Query Match 75.0%; Score 36; DB 2; Length 702;
  Best Local Similarity 88.9%; Pred. No. 43;
  Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 YLSGACLNL 9
|||||| |||
Db 605 YLSGANLNL 613

RESULT 2
T30835
breast cancer tumor suppressor Brca2 - mouse
C;Species: Mus musculus (house mouse)
C;Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 15-Mar-2004
C;Accession: T30835
R;Sharan, S.K.; Bradley, A.
Genomics 40, 234-241, 1997
A;Title: Murine Brca2: sequence, map position, and expression pattern.
A;Reference number: Z20894; MUID:97237041; PMID:9119389
A;Accession: T30835
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-3328 <SHA>
A;Cross-references: EMBL:U65594; NID:g1743859; PID:g1743860; PIDN:AAC23702.1
A;Experimental source: strain C57Bl/6
C;Genetics:
A;Gene: Brca2
A;Map position: 5
A;Note: expression of Brca2 detected in midgestation embryos and adult testis, thymus, ar
C;Superfamily: DNA recombination repair protein, BRCA2 type
C;Keywords: tumor suppressor

  Query Match 75.0%; Score 36; DB 2; Length 3328;
  Best Local Similarity 77.8%; Pred. No. 1.9e+02;
  Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 YLSGACLNL 9
||| ||||
Db 3011 YLSDECLNL 3019

RESULT 3
T42205
breast cancer susceptibility protein BRCA2 - mouse
C;Species: Mus musculus (house mouse)
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 15-Mar-2004
C;Accession: T42205
R;McAllister, K.A.; Haugen-Strano, A.; Hagevik, S.; Collins, N.K.; Brownlee, H.; Futreal,
submitted to the EMBL Data Library, February 1997
A;Description: Characterization of the mouse and rat homologs of the BRCA2 breast cancer
A;Reference number: Z22073
A;Accession: T42205
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-3329 <MCA>
A;Cross-references: EMBL:U89652; NID:g2443438; PID:g2443439; PIDN:AAB71377.1
A;Experimental source: strain CD1; 129Sv; ICR Swiss
C;Genetics:
A;Gene: BRCA2
C;Superfamily: DNA recombination repair protein, BRCA2 type

  Query Match 75.0%; Score 36; DB 2; Length 3329;
  Best Local Similarity 77.8%; Pred. No. 1.9e+02;
  Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 YLSGACLNL 9
||| ||||
Db 3011 YLSDECLNL 3019

RESULT 4
T30904
breast cancer tumor suppressor Brca2 - mouse
C;Species: Mus musculus (house mouse)
C;Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 09-Jul-2004
C;Accession: T30904
R;Connor, F.; Smith, A.; Wooster, R.; Stratton, M.; Dixon, A.; Campbell, E.; Tait, T.M.;
Hum. Mol. Genet. 6, 291-300, 1997
A;Title: Cloning, chromosomal mapping and expression pattern of the mouse Brca2 gene.
A;Reference number: Z20931; MUID:97217789; PMID:9063750

```
A;Accession: T30904
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-3329 <CON>
A;Cross-references: UNIPROT:P97929; EMBL:U82270; NID:g1854950; PID:g1854951; PIDN:AAB483
C;Genetics:
A;Gene: Brca2
A;Map position: 5
C;Superfamily: DNA recombination repair protein, BRCA2 type
C;Keywords: tumor suppressor

  Query Match             75.0%;  Score 36;  DB 2;  Length 3329;
  Best Local Similarity   77.8%;  Pred. No. 1.9e+02;
  Matches    7;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 YLSGACLNL 9
              |||  ||||
Db       3011 YLSDECLNL 3019


RESULT 5
S19088
dihydrolipoamide dehydrogenase (EC 1.8.1.4) - Enterococcus faecalis (fragment)
N;Alternate names: pyruvate dehydrogenase multienzyme complex chain E3
C;Species: Enterococcus faecalis
C;Date: 19-Mar-1997 #sequence_revision 06-Jun-1997 #text_change 30-Jun-1997
C;Accession: S19088
R;Allen, A.G.; Perham, R.N.
FEBS Lett. 287, 206-210, 1991
A;Title: Two lipoyl domains in the dihydrolipoamide acetyltransferase chain of the pyruv
A;Reference number: S16988; MUID:91348216; PMID:1908789
A;Accession: S19088
A;Molecule type: DNA
A;Residues: 1-57 <ALL>
A;Note: the source is designated as Streptococcus faecalis
C;Genetics:
A;Gene: pdhD
C;Superfamily: dihydrolipoamide dehydrogenase; dihydrolipoamide dehydrogenase homology
C;Keywords: FAD; flavoprotein; lipoamide; NAD; oxidoreductase; redox-active disulfide; t
F;11-39/Region: beta-alpha-beta FAD nucleotide-binding fold
F;47-52/Disulfide bonds: redox-active #status predicted

  Query Match             72.9%;  Score 35;  DB 2;  Length 57;
  Best Local Similarity   55.6%;  Pred. No. 6.5;
  Matches    5;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 YLSGACLNL 9
              |: | |||:
Db         42 YIGGVCLNV 50


RESULT 6
S56055
hypothetical protein YMR241w - yeast (Saccharomyces cerevisiae)
N;Alternate names: hypothetical protein YM9408.03
C;Species: Saccharomyces cerevisiae
C;Date: 27-Aug-1995 #sequence_revision 19-Oct-1995 #text_change 09-Jul-2004
C;Accession: S56055
R;Gentles, S.; Bowman, S.
submitted to the EMBL Data Library, March 1995
A;Reference number: S56053
A;Accession: S56055
A;Molecule type: DNA
A;Residues: 1-314 <GEN>
A;Cross-references: UNIPROT:Q04013; EMBL:Z48756; NID:g736304; PID:g736307; GSPDB:GN00013
C;Genetics:
A;Gene: SGD:YHM2; MIPS:YMR241w
A;Cross-references: SGD:S0004854
A;Map position: 13R
C;Superfamily: ADP,ATP carrier protein; ADP,ATP carrier protein repeat homology

  Query Match             72.9%;  Score 35;  DB 2;  Length 314;
  Best Local Similarity   87.5%;  Pred. No. 32;
  Matches    7;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          2 LSGACLNL 9
              | ||||||
Db         24 LLGACLNL 31


RESULT 7
A38971
polycystic kidney disease protein 1 precursor - human
C;Species: Homo sapiens (man)
C;Date: 26-Jan-1996 #sequence_revision 26-Jan-1996 #text_change 09-Jul-2004
C;Accession: A38971; A56520; A44604
R;Harris, P.C.
submitted to GenBank, May 1995
A;Reference number: A38971
A;Accession: A38971
A;Molecule type: mRNA
A;Residues: 1-4302 <HAR>
A;Cross-references: UNIPROT:P98161; GB:L33243; NID:g904222; PIDN:AAC37576.1; PID:g904223
R;Alexandra Gluecksmann-Kuis, M.; Tayber, O.; Woolf, E.A.; Bougueleret, L.; Deng, N.; Alp
 Z.; Torosian, S.; Zhou, J.; Reeders, S.T.; Bork, P.; Pohischmidt, M.; Loehning, C.; Krau
Cell 81, 289-298, 1995
A;Title: Polycystic kidney disease: the complete structure of the PKD1 gene and its prote
A;Reference number: A56520; MUID:95254638; PMID:7736581
A;Accession: A56520
A;Status: preliminary; nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 1-70,'E',72-137,'Q',139-252,'A',254-301,'D',303-690,'P',692-738,'R',740-762,'
5-1975,'V',1977-3389,'Q',3390-3980,'HV',3983-4003,'HV',4006-4302 <ALE>
A;Cross-references: GB:U24497; NID:g799334; PIDN:AAC50128.1; PID:g799335; GB:U24499
R;Ward, C.J.; Peral, B.; Hughes, J.; Thomas, S.; Gamble, V.; MacCarthy, A.B.; Sloane-Star
aris, J.J.; Dauwerse, H.G.; Peters, D.J.M.; Breuning, M.H.; Nellist, M.; Brook-Carter, P.
W.; van den Ouweland, A.M.W.; Eussen, B.; Verhoef, S.; Lindhout, D.; Halley, D.J.J.
Cell 77, 881-894, 1994
A;Title: The polycystic kidney disease 1 gene encodes a 14 kb transcript and lies within
A;Reference number: A44604; MUID:94273192; PMID:8004675
A;Accession: A44604
A;Status: significant sequence differences
A;Molecule type: mRNA
A;Cross-references: GB:L33243
R;Ward, C.J.; Peral, B.; Hughes, J.; Thomas, S.; Gamble, V.; MacCarthy, A.B.; Sloane-Star
aris, J.J.; Dauwerse, H.G.; Peters, D.J.M.; Breuning, M.H.; Nellist, M.; Brook-Carter, P.
W.; van den Ouweland, A.M.W.; Eussen, B.; Verhoef, S.; Lindhout, D.; Halley, D.J.J.
Cell 78, 724A, 1994
A;Reference number: A38972; MUID:94349375; PMID:8069919
A;Contents: annotation; erratum
A;Note: this is a revision to the sequence from reference A44604; the citation appears or
R;Ward, C.J.; Peral, B.; Hughes, J.; Thomas, S.; Gamble, V.; MacCarthy, A.B.; Sloane-Star
aris, J.J.; Dauwerse, H.G.; Peters, D.J.M.; Breuning, M.H.; Nellist, M.; Brook-Carter, P.
W.; van den Ouweland, A.M.W.; Eussen, B.; Verhoef, S.; Lindhout, D.; Halley, D.J.J.
Cell 81, 1171, 1995
A;Reference number: A56732
A;Contents: annotation; erratum
A;Note: this is a revision to the sequence from reference A44604
C;Genetics:
A;Gene: GDB:PKD1
A;Cross-references: GDB:120293; OMIM:173900; OMIM:601313
A;Map position: 16p13.3-16p13.3
C;Superfamily: human polycystic kidney disease protein 1; proteoglycan carboxyl-terminal
C;Keywords: duplication
F;1-23/Domain: signal sequence #status predicted <SIG>
F;24-4302/Product: polycystic kidney disease protein 1 #status predicted <MAT>
F;123-170/Domain: proteoglycan carboxyl-terminal homology <PCH>

  Query Match             72.9%;  Score 35;  DB 2;  Length 4302;
  Best Local Similarity   75.0%;  Pred. No. 3.7e+02;
  Matches    6;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 YLSGACLN 8
              || | |||
Db       2388 YLEGRCLN 2395
```

```
RESULT 8
I68703
MHC c6/g2 (Qa-2) protein - mouse (fragment)
C;Species: Mus musculus (house mouse)
C;Date: 02-Aug-1996 #sequence_revision 02-Aug-1996 #text_change 09-Jul-2004
C;Accession: I68703
R;Rogers, J.H.
Immunogenetics 21, 343-353, 1985
A;Title: Family organization of mouse H-2 class I genes.
A;Reference number: I54413; MUID:85206117; PMID:3997208
A;Accession: I68703
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-91 <RES>
A;Cross-references: UNIPROT:Q31147; GB:M14833; NID:g199298; PIDN:AAA39566.1; PID:g554214
C;Superfamily: class I histocompatibility antigen; immunoglobulin homology

  Query Match             70.8%;  Score 34;  DB 2;  Length 91;
  Best Local Similarity   85.7%;  Pred. No. 16;
  Matches    6;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 YLSGACL 7
              || ||||
Db         68 YLEGACL 74


RESULT 9
H84039
hypothetical protein BH3120 [imported] - Bacillus halodurans (strain C-125)
C;Species: Bacillus halodurans
C;Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 09-Jul-2004
C;Accession: H84039
R;Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fuji, F.; Hira
Nucleic Acids Res. 28, 4317-4331, 2000
A;Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and
A;Reference number: A83650; MUID:20512582; PMID:11058132
A;Accession: H84039
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-163 <STO>
A;Cross-references: UNIPROT:Q9K886; GB:AP001517; GB:BA000004; NID:g10175500; PIDN:BAB068
A;Experimental source: strain C-125
C;Genetics:
A;Gene: BH3120

  Query Match             70.8%;  Score 34;  DB 2;  Length 163;
  Best Local Similarity   75.0%;  Pred. No. 27;
  Matches    6;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 YLSGACLN 8
              || | |||
Db         80 YLIGGCLN 87


RESULT 10
B49181
beta B2-crystallin - chicken
C;Species: Gallus gallus (chicken)
C;Date: 19-Dec-1993 #sequence_revision 18-Nov-1994 #text_change 09-Jul-2004
C;Accession: B49181
R;Sawada, K.; Agata, K.; Eguchi, G.
Exp. Eye Res. 55, 879-887, 1992
A;Title: Crystallin gene expression in the process of lentoidogenesis in cultures of chi
A;Reference number: A49181; MUID:93137981; PMID:1283129
A;Accession: B49181
A;Status: preliminary
A;Molecule type: nucleic acid
A;Residues: 1-219 <SAW>
A;Cross-references: UNIPROT:Q05714; GB:S52930; NID:g264001; PIDN:AAB25042.1; PID:g264002
A;Experimental source: lens epithelial cells
A;Note: sequence extracted from NCBI backbone (NCBIN:123033, NCBIP:123034)
C;Superfamily: beta-crystallin
C;Keywords: duplication

  Query Match             70.8%;  Score 34;  DB 2;  Length 219;
  Best Local Similarity   87.5%;  Pred. No. 36;
  Matches    7;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          2 LSGACLNL 9
              |||| ||
Db         34 LSGACPNL 41


RESULT 11
H89772
hypothetical protein [imported] - Staphylococcus aureus (strain N315)
C;Species: Staphylococcus aureus
C;Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 09-Jul-2004
C;Accession: H89772
R;Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Oguch
ma, A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.; H
 C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.
Lancet 357, 1225-1240, 2001
A;Title: Whole genome sequencing of meticillin-resistant Stapylococcus aureus.
A;Reference number: A89758; MUID:21311952; PMID:11418146
A;Accession: H89772
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-254 <KUR>
A;Cross-references: UNIPROT:Q99X91; GB:BA000018; PID:g13700040; PIDN:BAB41339.1; GSPDB:GN
A;Experimental source: strain N315
C;Genetics:
A;Gene: SA0120

  Query Match             70.8%;  Score 34;  DB 2;  Length 254;
  Best Local Similarity   75.0%;  Pred. No. 41;
  Matches    6;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          2 LSGACLNL 9
              :|| ||||
Db        205 ISGRCLNL 212


RESULT 12
B84468
hypothetical protein At2g05410 [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 09-Jul-2004
C;Accession: B84468
R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.; M
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
Nature 402, 761-768, 1999
A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A;Reference number: A84420; MUID:20083487; PMID:10617197
A;Accession: B84468
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-265 <STO>
A;Cross-references: UNIPROT:Q9SHT3; GB:AE002093; NID:g4914359; PIDN:AAD32896.1; GSPDB:GN
C;Genetics:
A;Gene: At2g05410
A;Map position: 2

  Query Match             70.8%;  Score 34;  DB 2;  Length 265;
  Best Local Similarity   66.7%;  Pred. No. 43;
  Matches    6;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 YLSGACLNL 9
              ||  ||:||
Db        151 YLKTACMNL 159


RESULT 13
S15236
```

glutathione-disulfide reductase (EC 1.8.1.7) - Pseudomonas aeruginosa
C;Species: Pseudomonas aeruginosa
C;Date: 21-Nov-1993 #sequence_revision 10-Nov-1995 #text_change 09-Jul-2004
C;Accession: S15236; F83391
R;Perry, A.C.F.; Bhriain, N.N.; Brown, N.L.; Rouch, D.A.
Mol. Microbiol. 5, 163-171, 1991
A;Title: Molecular characterization of the gor gene encoding glutathione reductase from
ses.
A;Reference number: S15235; MUID:91194546; PMID:1849605
A;Accession: S15236
A;Molecule type: DNA
A;Residues: 1-451 <PER>
A;Cross-references: UNIPROT:P23189; EMBL:X54201; NID:g45324; PIDN:CAA38122.1; PID:g45326
R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrener, P.; Hickey, M.J.; Br
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,
.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A;Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
A;Reference number: A82950; MUID:20437337; PMID:10984043
A;Accession: F83391
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-451 <STO>
A;Cross-references: GB:AE004629; GB:AE004091; NID:g9948028; PIDN:AAG05413.1; GSPDB:GN001
A;Experimental source: strain PAO1
C;Genetics:
A;Gene: gor; PA2025
C;Superfamily: dihydrolipoamide dehydrogenase; dihydrolipoamide dehydrogenase homology
C;Keywords: FAD; flavoprotein; NADP; oxidoreductase; redox-active disulfide
F;8-443/Domain: dihydrolipoamide dehydrogenase homology <DLD>
F;42-47/Disulfide bonds: redox-active #status predicted

  Query Match 70.8%; Score 34; DB 2; Length 451;
  Best Local Similarity 55.6%; Pred. No. 70;
  Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 YLSGACLNL 9
|| | |:|:
Db 37 YLGGTCVNV 45

RESULT 14
S50820
surface protein type 51B - Paramecium tetraurelia
C;Species: Paramecium tetraurelia
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 09-Jul-2004
C;Accession: S50820
R;Scott, J.; Leeck, C.; Forney, J.
Nucleic Acids Res. 22, 5079-5084, 1994
A;Title: Analysis of the micronuclear B type surface protein gene in Paramecium tetraure
A;Reference number: S50820; MUID:95098630; PMID:7800503
A;Accession: S50820
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-2395 <SCO>
A;Cross-references: UNIPROT:Q27167; EMBL:U07603; NID:g467226; PIDN:AAA81947.1; PID:g4672
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, March 1994
C;Genetics:
A;Genetic code: SGC5
A;Introns: 472/3; 1310/3; 1821/3
C;Superfamily: G surface protein

  Query Match 70.8%; Score 34; DB 1; Length 2395;
  Best Local Similarity 100.0%; Pred. No. 3.3e+02;
  Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 YLSGAC 6
||||||
Db 892 YLSGAC 897

RESULT 15
T29869
hypothetical protein F10G2.9 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C;Accession: T29869
R;Murray, J.; Wohldmann, P.
submitted to the EMBL Data Library, July 1996
A;Description: The sequence of C. elegans cosmid F10G2.
A;Reference number: Z20701
A;Accession: T29869
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-114 <MUR>
A;Cross-references: UNIPROT:Q22960; EMBL:U64836; PIDN:AAB04834.1; GSPDB:GN00023; CESP:F1(
A;Experimental source: strain Bristol N2; clone F10G2
C;Genetics:
A;Gene: CESP:F10G2.9
A;Map position: 5
A;Introns: 61/3; 94/3

  Query Match 68.8%; Score 33; DB 2; Length 114;
  Best Local Similarity 55.6%; Pred. No. 30;
  Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 YLSGACLNL 9
|::| | ||
Db 47 YMAGNCFNL 55

Search completed: May 17, 2005, 06:20:05
Job time : 13.25 secs

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: May 17, 2005, 06:13:14 ; Search time 51.75 Seconds
(without alignments)
89.057 Million cell updates/sec

Title: US-10-725-373-5
Perfect score: 48
Sequence: 1 YLSGACLNL 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : UniProt_03:*
1: uniprot_sprot:*
2: uniprot_trembl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	39	81.2	270	2	Q90X37	Q90x37 brachydanio
2	39	81.2	285	2	Q8KN39	Q8kn39 sphingobium
3	37	77.1	445	2	Q8CMF3	Q8cmf3 streptococc
4	37	77.1	1864	2	Q6LFE6	Q6lfe6 plasmodium
5	37	77.1	5060	2	Q9K5M1	Q9k5m1 anabaena ci
6	36.5	76.0	430	2	Q9VED2	Q9ved2 drosophila
7	36	75.0	234	2	Q6WS99	Q6ws99 nocardia la
8	36	75.0	420	2	Q68DM9	Q68dm9 homo sapien
9	36	75.0	702	1	CEA5_HUMAN	P06731 homo sapien
10	36	75.0	702	2	Q8N4D0	Q8n4d0 homo sapien
11	36	75.0	1180	2	Q6PB43	Q6pb43 mus musculu
12	36	75.0	1624	2	Q9V3K8	Q9v3k8 drosophila
13	36	75.0	1637	2	Q95RU8	Q95ru8 drosophila
14	36	75.0	3329	1	BRC2_MOUSE	P97929 mus musculu
15	36	75.0	3329	2	Q8VHD0	Q8vhd0 mus musculu
16	35	72.9	286	2	Q88LI9	Q88li9 pseudomonas
17	35	72.9	294	2	Q642Y2	Q642y2 fugu rubrip
18	35	72.9	314	1	YHM2_YEAST	Q04013 saccharomyc
19	35	72.9	343	2	Q6A5T2	Q6a5t2 propionibac
20	35	72.9	437	2	Q8WPW6	Q8wpw6 paramecium
21	35	72.9	468	2	Q835M1	Q835m1 enterococcu
22	35	72.9	470	2	Q88VB6	Q88vb6 lactobacill
23	35	72.9	473	2	Q6N5V7	Q6n5v7 rhodopseudo
24	35	72.9	473	2	Q89KX2	Q89kx2 bradyrhizob
25	35	72.9	635	2	Q86J18	Q86j18 dictyosteli
26	35	72.9	682	2	Q6MC58	Q6mc58 parachlamyd
27	35	72.9	832	2	Q6XBY3	Q6xby3 acinetobact
28	35	72.9	1124	2	Q8ILS3	Q8ils3 plasmodium
29	35	72.9	3638	2	Q15142	Q15142 homo sapien
30	35	72.9	4303	1	PKD1_HUMAN	P98161 homo sapien
31	34	70.8	59	2	Q9KJD7	Q9kjd7 hafnia alve
32	34	70.8	91	2	Q31147	Q31147 mus musculu
33	34	70.8	144	2	Q6ZV57	Q6zv57 homo sapien
34	34	70.8	163	2	Q9K886	Q9k886 bacillus ha
35	34	70.8	205	2	Q6QIF5	Q6qif5 gallus gall
36	34	70.8	210	1	LOLA_CANBF	Q7vr38 candidatus
37	34	70.8	218	1	CRB2_CHICK	Q05714 gallus gall
38	34	70.8	219	2	Q6QIF4	Q6qif4 gallus gall
39	34	70.8	254	2	Q6X7T9	Q6x7t9 staphylococ
40	34	70.8	254	2	Q932K8	Q932k8 staphylococ
41	34	70.8	254	2	Q99X91	Q99x91 staphylococ
42	34	70.8	254	2	Q7A1Z3	Q7a1z3 staphylococ
43	34	70.8	254	2	Q6GD01	Q6gd01 staphylococ
44	34	70.8	254	2	Q6GKI1	Q6gki1 staphylococ
45	34	70.8	265	2	Q9SHT3	Q9sht3 arabidopsis

ALIGNMENTS

RESULT 1
Q90X37
ID Q90X37 PRELIMINARY; PRT; 270 AA.
AC Q90X37;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE SI:dZ129I22.1 (Novel gene similar to human and rodent IER5
DE (Immediately early response 5)).
GN Name=SI:dZ129I22.1;
OS Brachydanio rerio (Zebrafish) (Danio rerio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RA Grafham D.;
RL Submitted (JUN-2003) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AL592495; CAC95156.1; -.
DR Pfam; PF05760; IER; 1.
SQ SEQUENCE 270 AA; 29989 MW; 0B00F7756796E285 CRC64;

Query Match 81.2%; Score 39; DB 2; Length 270;
Best Local Similarity 87.5%; Pred. No. 40;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 YLSGACLN 8
| ||||||
Db 51 YYSGACLN 58

RESULT 2
Q8KN39
ID Q8KN39 PRELIMINARY; PRT; 285 AA.
AC Q8KN39;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Probable formyltetrahydrofolate deformylase.
OS Sphingobium chlorophenolicum.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Sphingomonadales;
OC Sphingomonadaceae; Sphingobium.
OX NCBI_TaxID=46429;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 39723;
RX MEDLINE=22159920; PubMed=12169590;
RX DOI=10.1128/JB.184.17.4672-4680.2002;
RA Cai M., Xun L.;
RT "Organization and regulation of pentachlorophenol-degrading genes in
RT Sphingobium chlorophenolicum ATCC 39723.";
RL J. Bacteriol. 184:4672-4680(2002).
DR EMBL; AF512952; AAM96665.1; -.

```
DR   GO; GO:0016597; F:amino acid binding; IEA.
DR   GO; GO:0008864; F:formyltetrahydrofolate deformylase activity; IEA.
DR   GO; GO:0016742; F:hydroxymethyl-, formyl- and related transfe. . .; IEA.
DR   GO; GO:0006189; P:'de novo' IMP biosynthesis; IEA.
DR   GO; GO:0009058; P:biosynthesis; IEA.
DR   InterPro; IPR002912; ACT.
DR   InterPro; IPR002110; ANK.
DR   InterPro; IPR002376; Formyl_transf_N.
DR   InterPro; IPR004810; PurU.
DR   Pfam; PF01842; ACT; 1.
DR   Pfam; PF00551; Formyl_trans_N; 1.
DR   PRINTS; PR01415; ANKYRIN.
DR   PRINTS; PR01575; FFH4HYDRLASE.
DR   TIGRFAMs; TIGR00655; PurU; 1.
SQ   SEQUENCE   285 AA;  31800 MW;  A29FB889683630F5 CRC64;

  Query Match             81.2%;  Score 39;  DB 2;  Length 285;
  Best Local Similarity   66.7%;  Pred. No. 41;
  Matches    6;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 YLSGACLNL 9
              |||| |:|:
Db        183 YLSGRCINI 191


RESULT 3
Q8CMF3
ID   Q8CMF3      PRELIMINARY;      PRT;   445 AA.
AC   Q8CMF3;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE   Hypothetical protein gbs0404 (Hypothetical protein gbs0734)
DE   (Hypothetical protein gbs0975).
GN   OrderedLocusNames=gbs0404, gbs0734, gbs0975;
OS   Streptococcus agalactiae (serotype III).
OC   Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC   Streptococcus.
OX   NCBI_TaxID=216495;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NEM316 / Serotype III;
RX   MEDLINE=22242508; PubMed=12354221;
RA   Glaser P., Rusniok C., Buchrieser C., Chevalier F., Frangeul L.,
RA   Msadek T., Zouine M., Couve E., Lalioui L., Poyart C., Trieu-Cuot P.,
RA   Kunst F.;
RT   "Genome sequence of Streptococcus agalactiae, a pathogen causing
RT   invasive neonatal disease.";
RL   Mol. Microbiol. 45:1499-1513(2002).
DR   EMBL; AL766845; CAD46048.1; -.
DR   EMBL; AL766846; CAD46378.1; -.
DR   EMBL; AL766848; CAD46634.1; -.
DR   SagaList; gbs0404; -.
DR   SagaList; gbs0734; -.
DR   SagaList; gbs0975; -.
DR   InterPro; IPR007921; CHAP.
DR   Pfam; PF05257; CHAP; 1.
DR   PROSITE; PS50911; CHAP; 1.
KW   Complete proteome; Hypothetical protein.
SQ   SEQUENCE   445 AA;  46940 MW;  3E3CD778F755C602 CRC64;

  Query Match             77.1%;  Score 37;  DB 2;  Length 445;
  Best Local Similarity   66.7%;  Pred. No. 1.5e+02;
  Matches    6;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 YLSGACLNL 9
              ||:| |:||
Db        319 YLAGQCVNL 327


RESULT 4
Q6LFE6
ID   Q6LFE6      PRELIMINARY;      PRT;  1864 AA.
AC   Q6LFE6;
DT   05-JUL-2004 (TrEMBLrel. 27, Created)
DT   05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT   05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE   Hypothetical protein.
GN   ORFNames=MAL6P1.303, PFF0595c;
OS   Plasmodium falciparum (isolate 3D7).
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=36329;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   PubMed=12368867; DOI=10.1038/nature01095;
RA   Hall N., Pain A., Berriman M., Churcher C., Harris B., Harris D.,
RA   Mungall K., Bowman S., Atkin R., Baker S., Barron A., Brooks K.,
RA   Buckee C.O., Burrows C., Cherevach I., Chillingworth C.,
RA   Chillingworth T., Christodoulou Z., Clark L., Clark R., Corto C.,
RA   Cronin A., Davies R., Davies P., Dear P., Dearden F., Doggett J.,
RA   Feltwell T., Goble A., Goodhead I., Gwilliam R., Hamlin N., Hance Z.,
RA   Harper D., Hauser H., Hornsby T., Holroyd S., Horrocks P.,
RA   Humphray S., Jagels K., James D., Johnson D., Kerhornou A., Knight A.,
RA   Kontfortov B., Keyes S., Larke N., Lawson D., Lennard N., Line A.,
RA   Maddison M., McLean J., Mooney P., Moule S., Murphy L., Oliver K.,
RA   Ormond D., Price C., Quail M.A., Rabbinowitsch E., Rajandream M-A.,
RA   Rutter S., Rutherford K.M., Sanders M., Simmonds M., Seeger K.,
RA   Sharp S., Smith R., Squares R., Squares S., Stevens K., Taylor K.,
RA   Tivey A., Unwin L., Whitehead S., Woodward J., Sulston J.E., Craig A.,
RA   Newbold C., Barrell B.G;
RT   "Sequence of Plasmodium falciparum chromosomes 1, 3-9 and 13.";
RL   Nature 419:527-531(2002).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Cherevach I., Davis P., Goodhead I., Stevens K., Mungall K.,
RA   Berriman M., Pain A., Hall N., Atkin R., Chillingworth C., Doggett J.,
RA   Ormond D., Sanders M., Hayes R., Hall S., Quail M., Barrell B.G;
RL   Submitted (MAR-2004) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; CR382399; CAG25362.1; -.
DR   InterPro; IPR001611; LRR.
DR   InterPro; IPR003885; LRR_cyst.
DR   Pfam; PF00560; LRR_1; 5.
DR   PRINTS; PR00019; LEURICHRPT.
DR   SMART; SM00365; LRR_SD22; 5.
KW   Hypothetical protein.
SQ   SEQUENCE   1864 AA;  220293 MW;  26221A5519868F85 CRC64;

  Query Match             77.1%;  Score 37;  DB 2;  Length 1864;
  Best Local Similarity   77.8%;  Pred. No. 5.2e+02;
  Matches    7;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 YLSGACLNL 9
              ||  |||||
Db       1488 YLKDACLNL 1496


RESULT 5
Q9K5M1
ID   Q9K5M1      PRELIMINARY;      PRT;  5060 AA.
AC   Q9K5M1;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Peptide synthetase.
GN   Name=adpB;
OS   Anabaena circinalis 90.
OC   Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Anabaena.
OX   NCBI_TaxID=46234;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=90;
RX   MEDLINE=20392447; PubMed=10931313;
RA   Rouhiainen L., Paulin L., Suomalainen S., Hyytiainen H., Buikema W.,
RA   Haselkorn R., Sivonen K.;
RT   "Genes encoding synthetases of cyclic depsipeptides,
RT   anabaenopeptilides, in Anabaena strain 90.";
```

```
RL   Mol. Microbiol. 37:156-167(2000).
CC   -!- SIMILARITY: Belongs to the ATP-dependent AMP-binding enzyme
CC       family.
DR   EMBL; AJ269505; CAC01604.1; -.
DR   HSSP; P14687; 1AMU.
DR   GO; GO:0003824; F:catalytic activity; IEA.
DR   GO; GO:0048037; F:cofactor binding; IEA.
DR   GO; GO:0008757; F:S-adenosylmethionine-dependent methyltransf. . .; IEA.
DR   GO; GO:0008152; P:metabolism; IEA.
DR   InterPro; IPR010071; AA_adenyl_dom.
DR   InterPro; IPR009081; ACP_like.
DR   InterPro; IPR000873; AMP-bind.
DR   InterPro; IPR001242; Condensatn.
DR   InterPro; IPR000276; GPCR_Rhodpsn.
DR   InterPro; IPR001601; Methyltransf.
DR   InterPro; IPR006163; Phsppanteth_bind.
DR   InterPro; IPR006162; Ppantne_S.
DR   InterPro; IPR000051; SAM_bind.
DR   Pfam; PF00501; AMP-binding; 4.
DR   Pfam; PF00668; Condensation; 4.
DR   Pfam; PF00550; PP-binding; 4.
DR   PRINTS; PR00154; AMPBINDING.
DR   TIGRFAMs; TIGR01733; AA-adenyl-dom; 4.
DR   PROSITE; PS50075; ACP_DOMAIN; 4.
DR   PROSITE; PS00455; AMP_BINDING; 4.
DR   PROSITE; PS00237; G_PROTEIN_RECEP_F1_1; UNKNOWN_3.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 2.
KW   Phosphopantetheine.
SQ   SEQUENCE   5060 AA;  565943 MW;  0A6D498ABC69093E CRC64;

  Query Match             77.1%;  Score 37;  DB 2;  Length 5060;
  Best Local Similarity   77.8%;  Pred. No. 1.2e+03;
  Matches    7;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 YLSGACLNL 9
              |:|||||  |
Db        682 YVSGACLYL 690


RESULT 6
Q9VED2
ID   Q9VED2      PRELIMINARY;      PRT;   430 AA.
AC   Q9VED2;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   CG14316-PA.
GN   ORFNames=CG14316;
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20196006; PubMed=10731132; DOI=10.1126/science.287.5461.2185;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.H., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,
RA   Abril J.F., Agbayani A., An H.J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., WoodageT, Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
RA   Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=22426065; PubMed=12537568;
RA   Celniker S.E., Wheeler D.A., Kronmiller B., Carlson J.W., Halpern A.,
RA   Patel S., Adams M., Champe M., Dugan S.P., Frise E., Hodgson A.,
RA   George R.A., Hoskins R.A., Laverty T., Muzny D.M., Nelson C.R.,
RA   Pacleb J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J.,
RA   Svirskas R., Tabor P.E., Wan K., Stapleton M., Sutton G.G., Venter C.,
RA   Weinstock G., Scherer S.E., Myers E.W., Gibbs R.A., Rubin G.M.;
RT   "Finishing a whole-genome shotgun: Release 3 of the Drosophila
RT   melanogaster euchromatic genome sequence.";
RL   Genome Biol. 3:RESEARCH0079-RESEARCH0079(2002).
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=22426070; PubMed=12537573;
RA   Kaminker J.S., Bergman C.M., Kronmiller B., Carlson J., Svirskas R.,
RA   Patel S., Frise E., Wheeler D.A., Lewis S.E., Rubin G.M.,
RA   Ashburner M., Celniker S.E.;
RT   "The transposable elements of the Drosophila melanogaster euchromatin:
RT   a genomics perspective.";
RL   Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).
RN   [4]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=22426069; PubMed=12537572;
RA   Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
RA   Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochnik S.E.,
RA   Smith C.D., Tupy J.L., Whitfied E.J., Bayraktaroglu L., Berman B.P.,
RA   Bettencourt B.R., Celniker S.E., de Grey A.D., Drysdale R.A.,
RA   Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,
RA   Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
RA   Lewis S.E.;
RT   "Annotation of the Drosophila melanogaster euchromatic genome: a
RT   systematic review.";
RL   Genome Biol. 3:RESEARCH0083-RESEARCH0083(2002).
RN   [5]
RP   SEQUENCE FROM N.A.
RG   FlyBase;
RL   Submitted (SEP-2002) to the EMBL/GenBank/DDBJ databases.
RN   [6]
RP   SEQUENCE FROM N.A.
RG   FlyBase;
RL   Submitted (MAR-2004) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AE003720; AAF55493.3; -.
DR   FlyBase; FBgn0038567; CG14316.
SQ   SEQUENCE   430 AA;  49293 MW;  EE0DB071C47C5CF2 CRC64;

  Query Match             76.0%;  Score 36.5;  DB 2;  Length 430;
  Best Local Similarity   75.0%;  Pred. No. 1.8e+02;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    3;  Gaps    1;

Qy          1 YLSGA---CLNL 9
              |||||   ||||
```

```
Db        191 YLSGAPLKCLNL 202


RESULT 7
Q6WS99
ID   Q6WS99      PRELIMINARY;      PRT;   234 AA.
AC   Q6WS99;
DT   05-JUL-2004 (TrEMBLrel. 27, Created)
DT   05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT   05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE   Putative NRPS (Fragment).
OS   Nocardia lactamdurans.
OC   Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC   Pseudonocardineae; Pseudonocardiaceae; Amycolatopsis.
OX   NCBI_TaxID=1913;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Ayuso A., Genilloud O.;
RL   Submitted (APR-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY271642; AAQ17095.1; -.
DR   GO; GO:0003824; F:catalytic activity; IEA.
DR   GO; GO:0008152; P:metabolism; IEA.
DR   InterPro; IPR000873; AMP-bind.
DR   PROSITE; PS00455; AMP_BINDING; 1.
FT   NON_TER       1      1
FT   NON_TER     234    234
SQ   SEQUENCE   234 AA;  24642 MW;  818369E1F82D2524 CRC64;

  Query Match             75.0%;  Score 36;  DB 2;  Length 234;
  Best Local Similarity   75.0%;  Pred. No. 1.3e+02;
  Matches    6;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 YLSGACLN 8
              ||:||||:
Db        195 YLAGACLS 202


RESULT 8
Q68DM9
ID   Q68DM9      PRELIMINARY;      PRT;   420 AA.
AC   Q68DM9;
DT   25-OCT-2004 (TrEMBLrel. 28, Created)
DT   25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT   25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE   Hypothetical protein DKFZp781M2392.
GN   Name=DKFZp781M2392;
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Colon carcinoma;
RG   The German cDNA Consortium;
RA   Poustka A., Albert R., Moosmayer P., Schupp I., Wellenreuther R.,
RA   Mewes H.W., Weil B., Amid C., Osanger A., Fobo G., Han M., Wiemann S.;
RL   Submitted (AUG-2004) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; CR749337; CAH18191.1; -.
DR   InterPro; IPR001589; Actbind_actnin.
DR   InterPro; IPR003599; Ig.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR003598; Ig_c2.
DR   Pfam; PF00047; ig; 3.
DR   SMART; SM00409; IG; 3.
DR   SMART; SM00408; IGc2; 3.
DR   PROSITE; PS00019; ACTININ_1; UNKNOWN_1.
DR   PROSITE; PS50835; IG_LIKE; 3.
KW   Hypothetical protein.
SQ   SEQUENCE   420 AA;  45508 MW;  6E30C0B4A00D0F59 CRC64;

  Query Match             75.0%;  Score 36;  DB 2;  Length 420;
  Best Local Similarity   88.9%;  Pred. No. 2.1e+02;
  Matches    8;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 YLSGACLNL 9
              ||||| |||
Db        323 YLSGANLNL 331


RESULT 9
CEA5_HUMAN
ID   CEA5_HUMAN     STANDARD;      PRT;   702 AA.
AC   P06731;
DT   01-JAN-1988 (Rel. 06, Created)
DT   01-DEC-1992 (Rel. 24, Last sequence update)
DT   25-OCT-2004 (Rel. 45, Last annotation update)
DE   Carcinoembryonic antigen-related cell adhesion molecule 5 precursor
DE   (Carcinoembryonic antigen) (CEA) (Meconium antigen 100) (CD66e
DE   antigen).
GN   Name=CEACAM5; Synonyms=CEA;
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=90258861; PubMed=2342461;
RA   Schrewe H., Thompson J., Bona M., Hefta L.J.F., Maruya A.,
RA   Hassauer M., Shively J.E., von Kleist S., Zimmermann W.;
RT   "Cloning of the complete gene for carcinoembryonic antigen: analysis
RT   of its promoter indicates a region conveying cell type-specific
RT   expression.";
RL   Mol. Cell. Biol. 10:2738-2748(1990).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=88038876; PubMed=3670312;
RA   Beauchemin N., Benchimol S., Cournoyer D., Fuks A., Stanners C.P.;
RT   "Isolation and characterization of full-length functional cDNA clones
RT   for human carcinoembryonic antigen.";
RL   Mol. Cell. Biol. 7:3221-3230(1987).
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=89122014; PubMed=3220478;
RA   Barnett T., Goebel S.J., Nothdurft M.A., Elting J.J.;
RT   "Carcinoembryonic antigen family: characterization of cDNAs coding for
RT   NCA and CEA and suggestion of nonrandom sequence variation in their
RT   conserved loop-domains.";
RL   Genomics 3:59-66(1988).
RN   [4]
RP   SEQUENCE OF 5-702 FROM N.A.
RX   MEDLINE=87128144; PubMed=3814146;
RA   Oikawa S., Nakazato H., Kosaki G.;
RT   "Primary structure of human carcinoembryonic antigen (CEA) deduced
RT   from cDNA sequence.";
RL   Biochem. Biophys. Res. Commun. 142:511-518(1987).
RN   [5]
RP   SEQUENCE OF 331-702 FROM N.A.
RX   MEDLINE=87204247; PubMed=3033671;
RA   Zimmermann W., Ortlieb B., Friedrich R., von Kleist S.;
RT   "Isolation and characterization of cDNA clones encoding the human
RT   carcinoembryonic antigen reveal a highly conserved repeating
RT   structure.";
RL   Proc. Natl. Acad. Sci. U.S.A. 84:2960-2964(1987).
CC   -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.
CC   -!- TISSUE SPECIFICITY: Found in adenocarcinomas of endodermally
CC       derived digestive system epithelium and fetal colon.
CC   -!- PTM: Complex immunoreactive glycoprotein with a MW of 180 kDa
CC       comprising 60% carbohydrate.
CC   -!- SIMILARITY: Belongs to the immunoglobulin superfamily. CEA family.
CC   -!- SIMILARITY: Contains 7 immunoglobulin-like domains.
CC   -!- DATABASE: NAME=PROW; NOTE=CD guide CD66e entry;
CC       WWW="http://www.ncbi.nlm.nih.gov/prow/cd/cd66e.htm".
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
```

```
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; M17303; AAB59513.1; -.
DR   EMBL; M59262; AAA62835.1; ALT_SEQ.
DR   EMBL; M59255; AAA62835.1; JOINED.
DR   EMBL; M59256; AAA62835.1; JOINED.
DR   EMBL; M59257; AAA62835.1; JOINED.
DR   EMBL; M59258; AAA62835.1; JOINED.
DR   EMBL; M59259; AAA62835.1; JOINED.
DR   EMBL; M59260; AAA62835.1; JOINED.
DR   EMBL; M59261; AAA62835.1; JOINED.
DR   EMBL; M59709; -; NOT_ANNOTATED_CDS.
DR   EMBL; M59710; -; NOT_ANNOTATED_CDS.
DR   EMBL; M29540; AAA51967.1; -.
DR   EMBL; X16455; CAA34474.1; -.
DR   EMBL; M15042; AAA51963.1; -.
DR   EMBL; M16234; AAA51972.1; -.
DR   PIR; A36319; A36319.
DR   PDB; 1E07; Model; A=35-676.
DR   Genew; HGNC:1817; CEACAM5.
DR   MIM; 114890; -.
DR   GO; GO:0005887; C:integral to plasma membrane; TAS.
DR   InterPro; IPR007110; Ig-like.
DR   Pfam; PF00047; ig; 6.
DR   PROSITE; PS50835; IG_LIKE; 6.
KW   3D-structure; Glycoprotein; GPI-anchor; Immunoglobulin domain;
KW   Lipoprotein; Membrane; Repeat; Signal.
FT   SIGNAL        1     34
FT   CHAIN        35    685       Carcinoembryonic antigen-related cell
FT                                 adhesion molecule 5.
FT   PROPEP      686    702       Removed in mature form (Potential).
FT   DOMAIN       35    144       Ig-like 1.
FT   DOMAIN      146    237       Ig-like 2.
FT   DOMAIN      238    322       Ig-like 3.
FT   DOMAIN      324    415       Ig-like 4.
FT   DOMAIN      416    498       Ig-like 5.
FT   DOMAIN      502    593       Ig-like 6.
FT   DOMAIN      594    677       Ig-like 7.
FT   CARBOHYD    104    104       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    115    115       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    152    152       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    182    182       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    197    197       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    204    204       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    208    208       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    246    246       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    256    256       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    274    274       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    288    288       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    292    292       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    309    309       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    330    330       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    351    351       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    360    360       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    375    375       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    432    432       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    466    466       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    480    480       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    508    508       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    529    529       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    553    553       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    560    560       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    580    580       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    612    612       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    650    650       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    665    665       N-linked (GlcNAc. . .) (Potential).
FT   LIPID       685    685       GPI-anchor amidated alanine (Potential).
FT   CONFLICT    320    320       Missing (in Ref. 4).
SQ   SEQUENCE   702 AA;  76795 MW;  6299AE26CDDBDB5C CRC64;
```

```
  Query Match           75.0%;  Score 36;  DB 1;  Length 702;
  Best Local Similarity  88.9%;  Pred. No. 3.4e+02;
  Matches    8;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 YLSGACLNL 9
              ||||| |||
Db        605 YLSGANLNL 613


RESULT 10
Q8N4D0
ID   Q8N4D0      PRELIMINARY;      PRT;   702 AA.
AC   Q8N4D0;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   CEACAM5 protein.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Colon;
RX   MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA   Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
RA   Jones S.J., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length human
RT   and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Colon;
RA   Strausberg R.;
RL   Submitted (JUL-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC034671; AAH34671.1; -.
DR   HSSP; Q61353; 1L6Z.
DR   InterPro; IPR001589; Actbind_actnin.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR003598; Ig_c2.
DR   Pfam; PF00047; ig; 6.
DR   SMART; SM00408; IGc2; 3.
DR   PROSITE; PS00019; ACTININ_1; UNKNOWN_3.
DR   PROSITE; PS50835; IG_LIKE; 6.
SQ   SEQUENCE   702 AA;  76781 MW;  97CCFB7399A0B05A CRC64;

  Query Match           75.0%;  Score 36;  DB 2;  Length 702;
  Best Local Similarity  88.9%;  Pred. No. 3.4e+02;
  Matches    8;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 YLSGACLNL 9
              ||||| |||
Db        605 YLSGANLNL 613


RESULT 11
Q6PB43
ID   Q6PB43      PRELIMINARY;      PRT;  1180 AA.
```

```
AC   Q6PB43;
DT   05-JUL-2004 (TrEMBLrel. 27, Created)
DT   05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT   05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE   Brca2 protein.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6; TISSUE=Brain;
RX   MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA   Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
RA   Jones S.J., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length human
RT   and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6; TISSUE=Brain;
RA   Strausberg R.;
RL   Submitted (OCT-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC059904; AAH59904.1; -.
DR   GO; GO:0005737; C:cytoplasm; IDA.
DR   GO; GO:0005634; C:nucleus; IDA.
DR   GO; GO:0005515; F:protein binding; IPI.
DR   GO; GO:0006974; P:response to DNA damage stimulus; TAS.
DR   InterPro; IPR011370; BRCA2.
DR   InterPro; IPR008994; Nucleic_acid_OB.
DR   PIRSF; PIRSF002397; BRCA2; 1.
SQ   SEQUENCE   1180 AA;  131684 MW;  C77CE050166EACAF CRC64;

  Query Match             75.0%;  Score 36;  DB 2;  Length 1180;
  Best Local Similarity   77.8%;  Pred. No. 5.3e+02;
  Matches    7;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 YLSGACLNL 9
              |||  |||| 
Db        862 YLSDECLNL 870


RESULT 12
Q9V3K8
ID   Q9V3K8      PRELIMINARY;      PRT;  1624 AA.
AC   Q9V3K8;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   EG:BACR37P7.2 protein.
GN   Name=EG:BACR37P7.2; ORFNames=CG2995;
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Madueno E., de Pablos B., Modolell J.;
RL   Submitted (MAY-1999) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Benos P.;
RL   Submitted (DEC-1999) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: Contains 1 SET domain.
DR   EMBL; AL050231; CAB65850.1; -.
DR   HSSP; Q8X225; 1PEG.
DR   FlyBase; FBgn0040372; CG2995.
DR   GO; GO:0005634; C:nucleus; IEA.
DR   GO; GO:0018024; F:histone-lysine N-methyltransferase activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0016568; P:chromatin modification; IEA.
DR   InterPro; IPR002110; ANK.
DR   InterPro; IPR007728; Pre-SET.
DR   InterPro; IPR001214; SET.
DR   InterPro; IPR003606; Zn2-binding.
DR   Pfam; PF00023; Ank; 5.
DR   Pfam; PF05033; Pre-SET; 1.
DR   Pfam; PF00856; SET; 1.
DR   PRINTS; PR01415; ANKYRIN.
DR   SMART; SM00248; ANK; 5.
DR   SMART; SM00468; PreSET; 1.
DR   SMART; SM00317; SET; 1.
DR   PROSITE; PS50088; ANK_REPEAT; 3.
DR   PROSITE; PS50297; ANK_REP_REGION; 1.
DR   PROSITE; PS50867; PRE_SET; 1.
DR   PROSITE; PS50280; SET; 1.
KW   ANK repeat.
SQ   SEQUENCE   1624 AA;  179964 MW;  AAA34BAFCDAE5121 CRC64;

  Query Match             75.0%;  Score 36;  DB 2;  Length 1624;
  Best Local Similarity   66.7%;  Pred. No. 7e+02;
  Matches    6;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 YLSGACLNL 9
              ||:| ||:|
Db       1084 YLNGTCLHL 1092


RESULT 13
Q95RU8
ID   Q95RU8      PRELIMINARY;      PRT;  1637 AA.
AC   Q95RU8;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE   LD10743p (CG2995-PA).
GN   Name=EG:BACR37P7.2; ORFNames=CG2995;
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Berkeley;
RA   Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA   Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
RA   Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
RA   Nunoo J., Pacleb J., Paragas V., Park S., Phouanenavong S., Wan K.,
RA   Yu C., Lewis S.E., Rubin G.M., Celniker S.;
RL   Submitted (OCT-2001) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20196006; PubMed=10731132; DOI=10.1126/science.287.5461.2185;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.H., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,
```

```
RA   Abril J.F., Agbayani A., An H.J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., WoodageT, Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
RA   Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=22426065; PubMed=12537568;
RA   Celniker S.E., Wheeler D.A., Kronmiller B., Carlson J.W., Halpern A.,
RA   Patel S., Adams M., Champe M., Dugan S.P., Frise E., Hodgson A.,
RA   George R.A., Hoskins R.A., Laverty T., Muzny D.M., Nelson C.R.,
RA   Pacleb J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J.,
RA   Svirskas R., Tabor P.E., Wan K., Stapleton M., Sutton G.G., Venter C.,
RA   Weinstock G., Scherer S.E., Myers E.W., Gibbs R.A., Rubin G.M.;
RT   "Finishing a whole-genome shotgun: Release 3 of the Drosophila
RT   melanogaster euchromatic genome sequence.";
RL   Genome Biol. 3:RESEARCH0079-RESEARCH0079(2002).
RN   [4]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=22426070; PubMed=12537573;
RA   Kaminker J.S., Bergman C.M., Kronmiller B., Carlson J., Svirskas R.,
RA   Patel S., Frise E., Wheeler D.A., Lewis S.E., Rubin G.M.,
RA   Ashburner M., Celniker S.E.;
RT   "The transposable elements of the Drosophila melanogaster euchromatin:
RT   a genomics perspective.";
RL   Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).
RN   [5]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=22426069; PubMed=12537572;
RA   Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
RA   Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochnik S.E.,
RA   Smith C.D., Tupy J.L., Whitfied E.J., Bayraktaroglu L., Berman B.P.,
RA   Bettencourt B.R., Celniker S.E., de Grey A.D., Drysdale R.A.,
RA   Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,
RA   Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
RA   Lewis S.E.;
RT   "Annotation of the Drosophila melanogaster euchromatic genome: a
RT   systematic review.";
RL   Genome Biol. 3:RESEARCH0083-RESEARCH0083(2002).
RN   [6]
RP   SEQUENCE FROM N.A.
RG   FlyBase;
RL   Submitted (SEP-2002) to the EMBL/GenBank/DDBJ databases.
RN   [7]
RP   SEQUENCE FROM N.A.
RG   FlyBase;
RL   Submitted (MAR-2004) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: Contains 1 SET domain.
DR   EMBL; AY061125; AAL28673.1; -.
DR   EMBL; AE003417; AAF45487.2; -.
DR   HSSP; Q8X225; 1PEG.
DR   IntAct; Q95RU8; -.
DR   FlyBase; FBgn0040372; CG2995.
DR   GO; GO:0005634; C:nucleus; IEA.
DR   GO; GO:0018024; F:histone-lysine N-methyltransferase activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0016568; P:chromatin modification; IEA.
DR   InterPro; IPR002110; ANK.
DR   InterPro; IPR007728; Pre-SET.
DR   InterPro; IPR001214; SET.
DR   InterPro; IPR003606; Zn2-binding.
DR   Pfam; PF00023; Ank; 6.
DR   Pfam; PF05033; Pre-SET; 1.
DR   Pfam; PF00856; SET; 1.
DR   PRINTS; PR01415; ANKYRIN.
DR   SMART; SM00248; ANK; 6.
DR   SMART; SM00468; PreSET; 1.
DR   SMART; SM00317; SET; 1.
DR   PROSITE; PS50088; ANK_REPEAT; 4.
DR   PROSITE; PS50297; ANK_REP_REGION; 1.
DR   PROSITE; PS50867; PRE_SET; 1.
DR   PROSITE; PS50280; SET; 1.
KW   ANK repeat.
SQ   SEQUENCE   1637 AA;  181241 MW;  A1DCC1356F1DDC88 CRC64;

  Query Match             75.0%;  Score 36;  DB 2;  Length 1637;
  Best Local Similarity   66.7%;  Pred. No. 7.1e+02;
  Matches    6;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 YLSGACLNL 9
              ||:| ||:|
Db       1084 YLNGTCLHL 1092


RESULT 14
BRC2_MOUSE
ID   BRC2_MOUSE     STANDARD;      PRT;  3329 AA.
AC   P97929; O35922; P97383;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   25-OCT-2004 (Rel. 45, Last annotation update)
DE   Breast cancer type 2 susceptibility protein homolog.
GN   Name=Brca2;
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=129;
RX   MEDLINE=97217789; PubMed=9063750; DOI=10.1093/hmg/6.2.291;
RA   Connor F., Smith A., Wooster R., Stratton M., Dixon A., Campbell E.,
RA   Tait T.M., Freeman T., Ashworth A.;
RT   "Cloning, chromosomal mapping and expression pattern of the mouse
RT   Brca2 gene.";
RL   Hum. Mol. Genet. 6:291-300(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6;
RX   MEDLINE=97237041; PubMed=9119389; DOI=10.1006/geno.1996.4573;
RA   Sharan S.K., Bradley A.;
RT   "Murine Brca2: sequence, map position, and expression pattern.";
RL   Genomics 40:234-241(1997).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=129/Sv;
RX   MEDLINE=97384941; PubMed=9242436;
RA   McAllister K.A., Haugen-Strano A., Hagevik S., Brownlee H.A.,
RA   Collins N.K., Futreal P.A., Bennett L.M., Wiseman R.W.;
```

```
RT   "Characterization of the rat and mouse homologues of the BRCA2 breast
RT   cancer susceptibility gene.";
RL   Cancer Res. 57:3121-3125(1997).
RN   [4]
RP   SEQUENCE OF 18-200 FROM N.A.
RX   MEDLINE=97075121; PubMed=8917547; DOI=10.1073/pnas.93.23.13078;
RA   Rajan J.V., Wang M., Marquis S.T., Chodosh L.A.;
RT   "Brca2 is coordinately regulated with Brca1 during proliferation and
RT   differentiation in mammary epithelial cells.";
RL   Proc. Natl. Acad. Sci. U.S.A. 93:13078-13083(1996).
RN   [5]
RP   SEQUENCE OF 569-625 FROM N.A.
RX   MEDLINE=97341126; PubMed=9196008;
RA   McAllister K.A., Ramachandran S., Haugen-Strano A., Fiedorek F.T. Jr.,
RA   Wiseman R.W.;
RT   "Genetic mapping of the Brca2 breast cancer susceptibility gene on
RT   mouse chromosome 5.";
RL   Mamm. Genome 8:540-541(1997).
CC   -!- FUNCTION: May participate in a pathway associated with the
CC       activation of double-strand break repair and/or homologous
CC       recombination (By similarity).
CC   -!- SUBUNIT: Interacts with RAD51 and DSS1.
CC   -!- TISSUE SPECIFICITY: Widely expressed. Highest expression in
CC       cerebellum, testis, ileum, appendix, epidymis, ovary and mammary
CC       gland. No expression in lung.
CC   -!- DEVELOPMENTAL STAGE: In the mammary gland, expression increases
CC       dramatically during pregnancy.
CC   -!- SIMILARITY: Contains 7 BRCA2 repeats.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U82270; AAB48306.1; -.
DR   EMBL; U72947; AAB40720.1; -.
DR   EMBL; U65594; AAC23702.1; -.
DR   EMBL; U89652; AAB71377.1; -.
DR   EMBL; U89503; AAC53276.1; -.
DR   PIR; T30904; T30904.
DR   PDB; 1MIU; X-ray; A=2378-3115.
DR   MGD; MGI:109337; Brca2.
DR   GO; GO:0005737; C:cytoplasm; IDA.
DR   GO; GO:0005634; C:nucleus; ISS.
DR   GO; GO:0005515; F:protein binding; ISS.
DR   GO; GO:0003697; F:single-stranded DNA binding; ISS.
DR   GO; GO:0006338; P:chromatin remodeling; ISS.
DR   GO; GO:0000724; P:double-strand break repair via homologous r. . .; ISS.
DR   GO; GO:0006325; P:establishment and/or maintenance of chromat. . .; ISS.
DR   GO; GO:0007093; P:mitotic checkpoint; ISS.
DR   GO; GO:0007090; P:regulation of S phase of mitotic cell cycle; ISS.
DR   GO; GO:0045449; P:regulation of transcription; ISS.
DR   InterPro; IPR002093; BRCA2_repeat.
DR   InterPro; IPR008994; Nucleic_acid_OB.
DR   Pfam; PF00634; BRCA2; 7.
DR   PIRSF; PIRSF002397; BRCA2; 1.
DR   PROSITE; PS50138; BRCA2_REPEAT; 6.
KW   3D-structure; DNA repair; Polymorphism; Repeat.
FT   REPEAT      981   1015       BRCA2 1.
FT   REPEAT     1192   1226       BRCA2 2.
FT   REPEAT     1394   1428       BRCA2 3.
FT   REPEAT     1491   1525       BRCA2 4.
FT   REPEAT     1623   1657       BRCA2 5.
FT   REPEAT     1924   1958       BRCA2 6.
FT   REPEAT     2004   2038       BRCA2 7.
FT   VARIANT      44     44       S -> F (in strain C57BL/6 and strain 129/
FT                                Sv).
FT   VARIANT     340    340       T -> P (in strain 129/Sv).
FT   VARIANT     377    377       N -> H (in strain C57BL/6).
FT   VARIANT     407    407       H -> P (in strain C57BL/6).
FT   VARIANT     661    661       I -> V (in strain C57BL/6).
FT   VARIANT     739    739       P -> H (in strain C57BL/6).
FT   VARIANT    1038   1038       I -> L (in strain C57BL/6 and strain 129/
FT                                Sv).
FT   VARIANT    1198   1199       GF -> RI (in strain C57BL/6).
FT   VARIANT    1257   1257       Q -> P (in strain C57BL/6).
FT   VARIANT    1392   1392       Q -> R (in strain C57BL/6).
FT   VARIANT    1520   1521       FD -> CG (in strain C57BL/6).
FT   VARIANT    1583   1583       R -> W (in strain C57BL/6).
FT   VARIANT    1613   1613       C -> W (in strain C57BL/6).
FT   VARIANT    1686   1686       S -> R (in strain C57BL/6).
FT   VARIANT    1799   1799       S -> F (in strain 129/Sv).
FT   VARIANT    1881   1881       P -> L (in strain C57BL/6).
FT   VARIANT    1894   1894       S -> F (in strain 129/Sv).
FT   VARIANT    2141   2141       Q -> K (in strain C57BL/6).
FT   VARIANT    2392   2392       S -> R (in strain C57BL/6).
FT   VARIANT    2605   2605       K -> Q (in strain C57BL/6).
FT   VARIANT    2648   2648       A -> P (in strain C57BL/6).
FT   VARIANT    2717   2717       R -> C (in strain 129/Sv).
FT   VARIANT    2729   2729       L -> M (in strain 129/Sv).
FT   VARIANT    2814   2814       Q -> H (in strain C57BL/6).
FT   VARIANT    2827   2827       A -> P (in strain C57BL/6).
FT   VARIANT    2907   2907       S -> I (in strain 129/Sv).
FT   VARIANT    2929   2929       H -> L (in strain 129/Sv).
FT   VARIANT    3058   3058       A -> G (in strain C57BL/6).
FT   VARIANT    3071   3071       A -> G (in strain C57BL/6).
FT   VARIANT    3081   3081       K -> E (in strain C57BL/6).
FT   VARIANT    3089   3089       T -> S (in strain C57BL/6).
FT   VARIANT    3105   3109       DSPKW -> SQSQV (in strain C57BL/6).
FT   VARIANT    3220   3220       A -> G (in strain 129/Sv).
FT   VARIANT    3238   3238       E -> K (in strain 129/Sv).
FT   VARIANT    3243   3243       Missing (in strain C57BL/6).
FT   VARIANT    3245   3245       R -> K (in strain 129/Sv).
FT   HELIX      2404   2420
FT   TURN       2421   2421
FT   HELIX      2430   2435
FT   TURN       2436   2436
FT   STRAND     2442   2443
FT   TURN       2444   2445
FT   HELIX      2446   2448
FT   TURN       2449   2449
FT   TURN       2457   2458
FT   TURN       2475   2480
FT   TURN       2485   2485
FT   HELIX      2486   2489
FT   TURN       2490   2492
FT   TURN       2494   2495
FT   TURN       2497   2498
FT   STRAND     2500   2501
FT   TURN       2503   2504
FT   STRAND     2507   2508
FT   TURN       2512   2513
FT   HELIX      2518   2527
FT   TURN       2529   2530
FT   TURN       2533   2534
FT   HELIX      2538   2554
FT   TURN       2555   2555
FT   TURN       2557   2558
FT   TURN       2560   2562
FT   HELIX      2570   2584
FT   TURN       2585   2586
FT   HELIX      2592   2597
FT   TURN       2598   2599
FT   STRAND     2604   2604
FT   STRAND     2607   2608
FT   STRAND     2610   2611
FT   STRAND     2642   2643
FT   STRAND     2648   2649
FT   HELIX      2655   2662
FT   TURN       2663   2664
FT   STRAND     2671   2671
FT   STRAND     2677   2677
FT   STRAND     2679   2680
```

```
FT   TURN       2688   2689
FT   STRAND     2696   2697
FT   HELIX      2701   2703
FT   STRAND     2707   2707
FT   STRAND     2714   2715
FT   HELIX      2726   2728
FT   STRAND     2736   2746
FT   STRAND     2750   2750
FT   STRAND     2753   2754
FT   STRAND     2760   2761
FT   TURN       2765   2766
FT   HELIX      2767   2778
FT   TURN       2779   2781
FT   HELIX      2782   2791
FT   TURN       2792   2793
FT   TURN       2811   2812
FT   HELIX      2813   2817
FT   TURN       2818   2819
FT   HELIX      2822   2830
FT   TURN       2836   2840
FT   TURN       2846   2847

  Query Match             75.0%;  Score 36;  DB 1;  Length 3329;
  Best Local Similarity   77.8%;  Pred. No. 1.3e+03;
  Matches    7;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 YLSGACLNL 9
              |||  ||||
Db       3011 YLSDECLNL 3019


RESULT 15
Q8VHD0
ID   Q8VHD0      PRELIMINARY;      PRT;  3329 AA.
AC   Q8VHD0;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Breast cancer 2.
GN   Name=Brca2;
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BALB/c; TISSUE=Testis;
RA   Callens N., Dumont M., Begue A., Van Lint C., Tranchant M., Samson C.,
RA   Tavtigian S.V., Baert J.-L., Simard J., de Launoit Y.;
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY033581; AAK57537.1; -.
DR   EMBL; AY033556; AAK57537.1; JOINED.
DR   EMBL; AY033557; AAK57537.1; JOINED.
DR   EMBL; AY033558; AAK57537.1; JOINED.
DR   EMBL; AY033559; AAK57537.1; JOINED.
DR   EMBL; AY033560; AAK57537.1; JOINED.
DR   EMBL; AY033561; AAK57537.1; JOINED.
DR   EMBL; AY033562; AAK57537.1; JOINED.
DR   EMBL; AY033563; AAK57537.1; JOINED.
DR   EMBL; AY033564; AAK57537.1; JOINED.
DR   EMBL; AY033565; AAK57537.1; JOINED.
DR   EMBL; AY033566; AAK57537.1; JOINED.
DR   EMBL; AY033567; AAK57537.1; JOINED.
DR   EMBL; AY033568; AAK57537.1; JOINED.
DR   EMBL; AY033569; AAK57537.1; JOINED.
DR   EMBL; AY033570; AAK57537.1; JOINED.
DR   EMBL; AY033571; AAK57537.1; JOINED.
DR   EMBL; AY033580; AAK57537.1; JOINED.
DR   EMBL; AY033579; AAK57537.1; JOINED.
DR   EMBL; AY033578; AAK57537.1; JOINED.
DR   EMBL; AY033577; AAK57537.1; JOINED.
DR   EMBL; AY033576; AAK57537.1; JOINED.
DR   EMBL; AY033575; AAK57537.1; JOINED.
DR   EMBL; AY033572; AAK57537.1; JOINED.
DR   EMBL; AY033573; AAK57537.1; JOINED.
DR   EMBL; AY033574; AAK57537.1; JOINED.
DR   HSSP; P97929; 1MIU.
DR   MGD; MGI:109337; Brca2.
DR   GO; GO:0005737; C:cytoplasm; IDA.
DR   GO; GO:0005634; C:nucleus; IDA.
DR   GO; GO:0005515; F:protein binding; IPI.
DR   GO; GO:0006974; P:response to DNA damage stimulus; TAS.
DR   Pfam; PF00634; BRCA2; 7.
DR   PIRSF; PIRSF002397; BRCA2; 1.
DR   PROSITE; PS50138; BRCA2_REPEAT; 6.
SQ   SEQUENCE   3329 AA;  370691 MW;  E14BA73F289E4FC5 CRC64;

  Query Match             75.0%;  Score 36;  DB 2;  Length 3329;
  Best Local Similarity   77.8%;  Pred. No. 1.3e+03;
  Matches    7;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 YLSGACLNL 9
              |||  ||||
Db       3011 YLSDECLNL 3019


Search completed: May 17, 2005, 06:23:41
Job time : 54.75 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: May 17, 2005, 06:13:14 ; Search time 66 Seconds
(without alignments)
52.740 Million cell updates/sec

Title: US-10-725-373-5
Perfect score: 48
Sequence: 1 YLSGACLNL 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_16Dec04:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	48	100.0	9	2	AAY09529	Aay09529 Carcinoem
2	38	79.2	502	7	ABO66668	Abo66668 Klebsiell
3	36.5	76.0	344	4	ABB69226	Abb69226 Drosophil
4	36	75.0	9	2	AAW39723	Aaw39723 Human car
5	36	75.0	9	2	AAW70045	Aaw70045 CEA deriv
6	36	75.0	9	2	AAW77134	Aaw77134 CEA synth
7	36	75.0	9	2	AAY47655	Aay47655 Immunogen
8	36	75.0	9	2	AAY09525	Aay09525 Carcinoem
9	36	75.0	9	2	AAY09526	Aay09526 Carcinoem
10	36	75.0	9	3	AAB13749	Aab13749 Peptide f
11	36	75.0	9	3	AAB13750	Aab13750 Peptide f
12	36	75.0	9	4	AAE02673	Aae02673 Human CEA
13	36	75.0	9	4	AAE00463	Aae00463 Human tum
14	36	75.0	9	4	AAB97818	Aab97818 Carcinoem
15	36	75.0	9	4	AAE05123	Aae05123 Carcinoem
16	36	75.0	9	4	AAE05124	Aae05124 Modified
17	36	75.0	9	4	AAB82776	Aab82776 Carcinoem
18	36	75.0	9	5	ABG79073	Abg79073 Human CEA
19	36	75.0	9	5	AAE26805	Aae26805 Human HLA
20	36	75.0	9	5	AAU95893	Aau95893 Immunogen
21	36	75.0	9	5	AAB47917	Aab47917 Modified
22	36	75.0	9	5	AAE19089	Aae19089 HLA-A24 r
23	36	75.0	9	5	AAE19088	Aae19088 HLA-A24 r
24	36	75.0	9	6	ABR56428	Abr56428 CEA epito
25	36	75.0	9	6	ABP98779	Abp98779 CAE pepti
26	36	75.0	9	6	ABR44529	Abr44529 CEA epito
27	36	75.0	9	7	ADD84715	Add84715 Human car
28	36	75.0	9	7	AAO24210	Aao24210 Human tum
29	36	75.0	9	8	ADG20333	Adg20333 Antigenic
30	36	75.0	9	8	ADJ36382	Adj36382 CEA epito
31	36	75.0	9	8	ADM12344	Adm12344 MHC class
32	36	75.0	9	8	ADM12341	Adm12341 MHC class
33	36	75.0	9	8	ADM72999	Adm72999 Human CEA
34	36	75.0	9	8	ADL46188	Adl46188 Human CAP
35	36	75.0	9	8	ADN63713	Adn63713 HLA bindi
36	36	75.0	9	8	ADO38561	Ado38561 Carcinoem
37	36	75.0	9	8	ADO38564	Ado38564 Carcinoem
38	36	75.0	10	2	AAY46555	Aay46555 Immunogen
39	36	75.0	10	5	AAU11587	Aau11587 Human car
40	36	75.0	10	6	ABR83489	Abr83489 Human car
41	36	75.0	10	8	ADM72998	Adm72998 Human CEA
42	36	75.0	10	8	ADP80031	Adp80031 Human HLA
43	36	75.0	14	4	AAB88124	Aab88124 CD66 pept
44	36	75.0	25	5	AAU82083	Aau82083 T-cell sp
45	36	75.0	27	5	AAU82075	Aau82075 T-cell sp

ALIGNMENTS

RESULT 1
AAY09529
ID AAY09529 standard; peptide; 9 AA.
XX
AC AAY09529;
XX
DT 20-JUL-1999 (first entry)
XX
DE Carcinoembryonic antigen peptide agonist SEQ ID NO:5.
XX
KW Carcinoembryonic antigen; CEA; human; agonist; antagonist;
KW immune response; carcinoma; gastrointestinal; breast; pancreatic;
KW bladder; ovarian; lung; prostatic; T cell proliferation; cancer;
KW adoptive transfer therapy; autoimmune reaction; immunotherapy.
XX
OS Homo sapiens.
OS Synthetic.
XX
PN WO9919478-A1.
XX
PD 22-APR-1999.
XX
PF 22-SEP-1998; 98WO-US019794.
XX
PR 10-OCT-1997; 97US-0061589P.
XX
PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
XX
PI Schlom J, Barzaga E, Zaremba S;
XX
DR WPI; 1999-326544/27.
XX
PT Peptide agonists and antagonists of carcinoembryonal antigen.
XX
PS Claim 5; Page 53; 72pp; English.
XX
CC The present invention describes peptides (A) that comprise agonists (Ia)
CC or antagonists (Ib) of human carcinoembryonal antigen (CEA). (Ia) are
CC used in vaccines to kill or inhibit carcinoma cells that express CEA or
CC its epitopes, particularly for treating gastrointestinal, breast,
CC pancreatic, bladder, ovarian, lung or prostatic carcinoma. They can also
CC be used to proliferate T cells, e.g. from vaccinated subjects, for use in
CC adoptive transfer therapy. (Ib) are used to inhibit CEA-specific immune
CC responses, e.g. in vaccinated subjects, to prevent an autoimmune reaction
CC to cancer immunotherapy (i.e. to prevent attack on normal but CEA-
CC expressing cells). (Ia) are more active than native sequence (I) and
CC generate a highly specific and systemic anti-CEA response. Cytotoxic T
CC cells generated recognize both (Ia) and native CEA epitopes. The present

```
CC   sequence represents a specifically claimed example of (Ia)
XX
SQ   Sequence 9 AA;

  Query Match             100.0%;  Score 48;  DB 2;  Length 9;
  Best Local Similarity   100.0%;  Pred. No. 1.8e+06;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 YLSGACLNL 9
              |||||||||
Db          1 YLSGACLNL 9


RESULT 2
ABO66668
ID   ABO66668 standard; protein; 502 AA.
XX
AC   ABO66668;
XX
DT   29-JUL-2004  (first entry)
XX
DE   Klebsiella pneumoniae polypeptide seqid 13185.
XX
KW   Recombinant expression vector; transcription regulatory element;
KW   Klebsiella pneumoniae protein; antibacterial; Vaccine.
XX
OS   Klebsiella pneumoniae.
XX
PN   US6610836-B1.
XX
PD   26-AUG-2003.
XX
PF   27-JAN-2000; 2000US-00489039.
XX
PR   29-JAN-1999;   99US-0117747P.
XX
PA   (GENO-) GENOME THERAPEUTICS CORP.
XX
PI   Breton GL,  Osborne M;
XX
DR   WPI; 2003-895346/82.
DR   N-PSDB; ABD00239.
XX
PT   New nucleic acid encoding a Klebsiella pneumoniae polypeptide, useful for
PT   preparing a vaccine composition against Klebsiella pneumoniae.
XX
PS   Disclosure; SEQ ID NO 13185; 932pp; English.
XX
CC   The invention describes a new isolated nucleic acid encoding a Klebsiella
CC   pneumoniae polypeptide. Also described are: a recombinant expression
CC   vector comprising the nucleic acid, operably linked to a transcription
CC   regulatory element; and a cell comprising the recombinant expression
CC   vector. The nucleic acid is useful for preparing a vaccine composition
CC   against Klebsiella pneumoniae. This is the amino acid sequence of a
CC   Klebsiella pneumoniae polypeptide of the invention
XX
SQ   Sequence 502 AA;

  Query Match             79.2%;  Score 38;  DB 7;  Length 502;
  Best Local Similarity   75.0%;  Pred. No. 1.1e+02;
  Matches    6;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 YLSGACLN 8
              ::||||||
Db        401 FISGACLN 408


RESULT 3
ABB69226
ID   ABB69226 standard; protein; 344 AA.
XX
AC   ABB69226;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster polypeptide SEQ ID NO 34470.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US009231.
XX
PR   23-MAR-2000; 2000US-0191637P.
PR   11-JUL-2000; 2000US-00614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   N-PSDB; ABL13329.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signaling and cell-cell
PT   interactions.
XX
PS   Disclosure; SEQ ID NO 34470; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
CC   ABB72072). The sequence data for this patent did not form part of the
CC   printed specification, but was obtained in electronic format directly
CC   from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 344 AA;

  Query Match             76.0%;  Score 36.5;  DB 4;  Length 344;
  Best Local Similarity   75.0%;  Pred. No. 1.4e+02;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    3;  Gaps    1;

Qy          1 YLSGA---CLNL 9
              |||||   ||||
Db        134 YLSGAPLKCLNL 145


RESULT 4
AAW39723
ID   AAW39723 standard; peptide; 9 AA.
XX
AC   AAW39723;
XX
DT   11-JUN-1998  (first entry)
XX
DE   Human carcina-embryonic antigen (CEA) peptide (pos. 571-579).
XX
KW   T cell epitope; immune response; human leukocyte antigen; HLA Class I;
KW   vaccine; immunogenic; major histocompatablilty complex; MHC; B cell;
KW   disease; anti-tumour; anti-viral.
XX
OS   Homo sapiens.
XX
PN   WO9741440-A1.
XX
PD   06-NOV-1997.
XX
```

PF 28-APR-1997; 97WO-NL000229.
XX
PR 26-APR-1996; 96EP-00201145.
PR 23-DEC-1996; 96EP-00203670.
XX
PA (UYLE-) RIJKSUNIV LEIDEN.
PA (SCIS-) SCI SEED CAPITAL INVESTMENTS BV.
XX
PI Van Der Burg SH, Kast WM, Toes REM, Offringa R, Melief CJM;
XX
DR WPI; 1997-549891/50.
XX
PT Method of selecting T cell peptide epitope(s) - by measuring the
PT stability of HLA class I-peptide complexes on intact B cells.
XX
PS Example 3; Page 85; 109pp; English.
XX
CC Peptides AAW39430-W39734 are used in a novel method for the selection of
CC immunogenic T-cell peptide epitopes present in polypeptide antigens. The
CC method involves the identification of peptide sequences capable of
CC binding to an HLA (human leukocyte antigen) class I molecule and
CC measuring the binding of this epitope peptide to the HLA class I peptide.
CC The stability of binding of the peptide and MHC (major histocompatablilty
CC complex) class I molecule is measured on intact human B cells carrying
CC the MHC molecule at their cell surfaces. The method can be used to select
CC peptide epitopes for generating vaccines against a disease associated
CC with the polypeptide, e.g. cancers or AIDS. The peptide epitopes are
CC especially T-cell peptide epitopes with strong anti-tumour and anti-viral
CC immune responses. Peptide AAW39723 is derived from the human carcino-
CC embryonic antigen (CEA) and has the ability to bind to the human MHC
CC Class I allele HLA-A2.1
XX
SQ Sequence 9 AA;

Query Match 75.0%; Score 36; DB 2; Length 9;
Best Local Similarity 88.9%; Pred. No. 1.8e+06;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 YLSGACLNL 9
||||| |||
Db 1 YLSGANLNL 9

RESULT 5
AAW70045
ID AAW70045 standard; peptide; 9 AA.
XX
AC AAW70045;
XX
DT 22-OCT-1998 (first entry)
XX
DE CEA derived HLA-A2.1 binding peptide 2 (residues 605-613).
XX
KW Cytotoxic T lymphocyte; CTL; major histocompatibility complex; MHC;
KW human leukocyte antigen; HLA; tumour associated antigen; cancer;
KW antigen presenting cell; APC; immunogenic peptide; immune disorder;
KW viral infection; AIDS; hepatitis; bacterial infection; malaria; CEA;
KW fungal infection; tuberculosis; melanoma; carcinoembryonic antigen.
XX
OS Synthetic.
OS Homo sapiens.
XX
PN WO9833888-A1.
XX
PD 06-AUG-1998.
XX
PF 30-JAN-1998; 98WO-US001959.
XX
PR 31-JAN-1997; 97US-0036696P.
XX
PA (EPIM-) EPIMMUNE INC.
XX
PI Tsai V, Southwood S, Sidney J, Sette A, Celis E;
XX
DR WPI; 1998-437445/37.
XX
PT Production of antigen-specific cytotoxic T cells - by incubating
PT immunogenic peptide(s) from antigen that binds class I major
PT histocompatibility complex molecules with pre-treated antigen presenting
PT cells.
XX
PS Example 6; Page 75; 104pp; English.
XX
CC Sequences shown in AAW70044 to AAW70052 represent peptides derived from
CC carcinoembryonic antigen (CEA). The peptides can bind to a human
CC leukocyte antigen (HLA), HLA-A2.1 and are used to exemplify the method of
CC invention of producing antigen-specific cytotoxic T cells (CTLs) in
CC vitro. The method comprises contacting immunogenic peptides from an
CC antigen that binds class I major histocompatibility complex (MHC)
CC molecules with antigen presenting cells (APCs) pretreated with
CC pretreatment growth factors, and incubating the APCs with purified CD8
CC cells in the presence of at least 2 incubation growth factors, thereby
CC producing antigen-specific CTLs. A method for specifically killing target
CC cells in a human patient is also provided which comprises obtaining a
CC fluid sample containing CTLs from a patient, contacting the cytotoxic T
CC cells with APCs pretreated with pre-treatment growth factors, where the
CC APCs comprise class I MHC molecules. The pretreated APCs are incubated
CC with the cytotoxic growth factors, thereby producing activated CTLs which
CC are contacted with a carrier to form a composition. The composition can
CC then be administered to the patient. The activated CTLs can be used for
CC treating cancers, immune disorders, viral infections, AIDS, hepatitis,
CC bacterial infection, fungal infection, malaria or tuberculosis
XX
SQ Sequence 9 AA;

Query Match 75.0%; Score 36; DB 2; Length 9;
Best Local Similarity 88.9%; Pred. No. 1.8e+06;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 YLSGACLNL 9
||||| |||
Db 1 YLSGANLNL 9

RESULT 6
AAW77134
ID AAW77134 standard; peptide; 9 AA.
XX
AC AAW77134;
XX
DT 16-NOV-1998 (first entry)
XX
DE CEA synthetic peptide epitope 1.
XX
KW Tyrosinase; tyrosinase cytotoxic lymphocyte response;
KW cytotoxic T lymphocyte; cysteine-depleted; melanoma.
XX
OS Synthetic.
XX
PN WO9833810-A2.
XX
PD 06-AUG-1998.
XX
PF 29-JAN-1998; 98WO-US001592.
XX
PR 30-JAN-1997; 97US-0037781P.
XX
PA (UYVI-) UNIV VIRGINIA PATENT FOUND.
XX
PI Slingluff CL, Hunt DF, Engelhard VH, Kittlesen D;
XX
DR WPI; 1998-437388/37.
XX
PT Disease specific immunogen - comprises disease specific cytotoxic T
PT lymphocyte epitope used to elicit melanoma specific CTL response.
XX

PS Disclosure; Page 27; 93pp; English.
XX
CC The peptide epitope AAW77119-W77138 were created for human tumour-
CC specific cytotoxic T lymphocyte response. These peptides are are cysteine
CC - depleted mutants of a native disease-specific CTL epitope. The cysteine
CC - depleted CTL epitopes elicit a stronger or more specific CTL response
CC than the native epitope. The epitopes can be used in a disease-specific
CC immunogen to protect a mammal against disease in particular melanomas.
CC The peptides may also be used to screen a sample for the presence of an
CC antigen with the same epitope, or with a different cross-reactive epitope
XX
SQ Sequence 9 AA;

Query Match 75.0%; Score 36; DB 2; Length 9;
Best Local Similarity 88.9%; Pred. No. 1.8e+06;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 YLSGACLNL 9
|||||.|||
Db 1 YLSGANLNL 9

RESULT 7
AAY47655
ID AAY47655 standard; peptide; 9 AA.
XX
AC AAY47655;
XX
DT 01-DEC-1999 (first entry)
XX
DE Immunogenic peptide having a human leukocyte antigen binding motif #2266.
XX
KW Human leukocyte antigen; binding; immunogenic; glycoprotein; MHC; HLA;
KW immune response; T cell activation; major histocompatibility complex;
KW cytotoxic T lymphocyte; CTL; tumour rejection; viral infection; cancer;
KW prostate cancer; hepatitis B; hepatitis C; AIDS; renal carcinoma;
KW vaccine; immunisation.
XX
OS Synthetic.
OS Homo sapiens.
XX
PN WO9945954-A1.
XX
PD 16-SEP-1999.
XX
PF 13-MAR-1998; 98WO-US005039.
XX
PR 13-MAR-1998; 98WO-US005039.
XX
PA (EPIM-) EPIMMUNE INC.
XX
PI Sette A, Kubo RT, Sidney J, Celis E, Grey HM, Southwood S;
XX
DR WPI; 1999-551214/46.
XX
PT New immunogenic peptides with HLA binding motif, useful in treatment and
PT diagnosis of cancers and viral diseases.
XX
PS Claim 1; Page 118; 150pp; English.
XX
CC AAY45390 to AAY48214 represent specifically claimed immunogenic peptides
CC having a human major histocompatibility complex (MHC) Class I (also known
CC as human leukocyte antigen (HLA)) binding motif. The immunogenic peptides
CC can bind to a specific HLA allele (i.e. HLA-A subtypes HLA-A2.1, A1, A3.2
CC or A24.1 or HLA-B or C) and induce a cytotoxic T cell response against
CC the antigen from which the peptide is derived. Cytotoxic T lymphocytes
CC (CTLs) which destroy antigen-bearing cells are normally induced by an
CC antigen in the form of a peptide fragment bound to a HLA molecule, rather
CC than the intact foreign antigen itself, and are particularly important in
CC tumour rejection and in fighting viral infections. The peptides are
CC therefore useful therapeutically to treat or prevent viral infections and
CC cancers in mammals (especially humans) e.g. prostate cancer, hepatitis B
CC and C, AIDS, and renal carcinoma. They can be administered as vaccines to
CC elicit an immune response in individuals susceptible or otherwise at risk
CC of viral infection or cancer, or used to treat chronic or acute
CC conditions. They are also useful diagnostically, and can be used to
CC induce a cytotoxic T cell response, by contacting a cytotoxic T cell with
CC the peptide e.g. to produce CTLs ex vivo for infusion back into a
CC patient. The polynucleotides encoding the immunogenic peptides are also
CC useful therapeutically and for immunisation as above
XX
SQ Sequence 9 AA;

Query Match 75.0%; Score 36; DB 2; Length 9;
Best Local Similarity 88.9%; Pred. No. 1.8e+06;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 YLSGACLNL 9
|||||.|||
Db 1 YLSGANLNL 9

RESULT 8
AAY09525
ID AAY09525 standard; peptide; 9 AA.
XX
AC AAY09525;
XX
DT 20-JUL-1999 (first entry)
XX
DE Carcinoembryonic antigen peptide agonist CAP-1.
XX
KW Carcinoembryonic antigen; CEA; human; agonist; antagonist;
KW immune response; carcinoma; gastrointestinal; breast; pancreatic;
KW bladder; ovarian; lung; prostatic; T cell proliferation; cancer;
KW adoptive transfer therapy; autoimmune reaction; immunotherapy.
XX
OS Homo sapiens.
OS Synthetic.
XX
PN WO9919478-A1.
XX
PD 22-APR-1999.
XX
PF 22-SEP-1998; 98WO-US019794.
XX
PR 10-OCT-1997; 97US-0061589P.
XX
PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
XX
PI Schlom J, Barzaga E, Zaremba S;
XX
DR WPI; 1999-326544/27.
XX
PT Peptide agonists and antagonists of carcinoembryonal antigen.
XX
PS Claim 1; Page 53; 72pp; English.
XX
CC The present invention describes peptides (A) that comprise agonists (Ia)
CC or antagonists (Ib) of human carcinoembryonal antigen (CEA). (Ia) are
CC used in vaccines to kill or inhibit carcinoma cells that express CEA or
CC its epitopes, particularly for treating gastrointestinal, breast,
CC pancreatic, bladder, ovarian, lung or prostatic carcinoma. They can also
CC be used to proliferate T cells, e.g. from vaccinated subjects, for use in
CC adoptive transfer therapy. (Ib) are used to inhibit CEA-specific immune
CC responses, e.g. in vaccinated subjects, to prevent an autoimmune reaction
CC to cancer immunotherapy (i.e. to prevent attack on normal but CEA-
CC expressing cells). (Ia) are more active than native sequence (I) and
CC generate a highly specific and systemic anti-CEA response. Cytotoxic T
CC cells generated recognize both (Ia) and native CEA epitopes. The present
CC sequence represents a specifically claimed example of (Ia)
XX
SQ Sequence 9 AA;

Query Match 75.0%; Score 36; DB 2; Length 9;
Best Local Similarity 88.9%; Pred. No. 1.8e+06;

Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy          1 YLSGACLNL 9
              ||||| |||
Db          1 YLSGANLNL 9
```

RESULT 9
AAY09526
ID AAY09526 standard; peptide; 9 AA.
XX
AC AAY09526;
XX
DT 20-JUL-1999 (first entry)
XX
DE Carcinoembryonic antigen peptide agonist SEQ ID NO:2.
XX
KW Carcinoembryonic antigen; CEA; human; agonist; antagonist;
KW immune response; carcinoma; gastrointestinal; breast; pancreatic;
KW bladder; ovarian; lung; prostatic; T cell proliferation; cancer;
KW adoptive transfer therapy; autoimmune reaction; immunotherapy.
XX
OS Homo sapiens.
OS Synthetic.
XX
PN WO9919478-A1.
XX
PD 22-APR-1999.
XX
PF 22-SEP-1998; 98WO-US019794.
XX
PR 10-OCT-1997; 97US-0061589P.
XX
PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
XX
PI Schlom J, Barzaga E, Zaremba S;
XX
DR WPI; 1999-326544/27.
XX
PT Peptide agonists and antagonists of carcinoembryonal antigen.
XX
PS Claim 5; Page 53; 72pp; English.
XX
CC The present invention describes peptides (A) that comprise agonists (Ia)
CC or antagonists (Ib) of human carcinoembryonal antigen (CEA). (Ia) are
CC used in vaccines to kill or inhibit carcinoma cells that express CEA or
CC its epitopes, particularly for treating gastrointestinal, breast,
CC pancreatic, bladder, ovarian, lung or prostatic carcinoma. They can also
CC be used to proliferate T cells, e.g. from vaccinated subjects, for use in
CC adoptive transfer therapy. (Ib) are used to inhibit CEA-specific immune
CC responses, e.g. in vaccinated subjects, to prevent an autoimmune reaction
CC to cancer immunotherapy (i.e. to prevent attack on normal but CEA-
CC expressing cells). (Ia) are more active than native sequence (I) and
CC generate a highly specific and systemic anti-CEA response. Cytotoxic T
CC cells generated recognize both (Ia) and native CEA epitopes. The present
CC sequence represents a specifically claimed example of (Ia)
XX
SQ Sequence 9 AA;

Query Match 75.0%; Score 36; DB 2; Length 9;
Best Local Similarity 88.9%; Pred. No. 1.8e+06;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy          1 YLSGACLNL 9
              ||||| |||
Db          1 YLSGADLNL 9
```

RESULT 10
AAB13749
ID AAB13749 standard; peptide; 9 AA.
XX
AC AAB13749;
XX
DT 02-FEB-2001 (first entry)
XX
DE Peptide fragment # 1 from human CEA.
XX
KW Human; T-cell; immune response; antigen; epitope; B7 family molecule;
KW Leukocyte function-associated antigen-3; LFA-3;
KW Intercellular adhesion molecule-1; ICAM-1; vaccine; immunotherapy;
KW colon polyp; Crohn's disease; ulcerative colitis; breast lesion; tumour;
KW CEA.
XX
OS Homo sapiens.
XX
PN WO200034494-A1.
XX
PD 15-JUN-2000.
XX
PF 12-NOV-1999; 99WO-US026866.
XX
PR 09-DEC-1998; 98US-0111582P.
XX
PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
PA (THER-) THERION BIOLOGICS CORP.
XX
PI Schlom J, Hodge J, Panicali D;
XX
DR WPI; 2000-431307/37.
XX
PT Novel recombinant vector useful as immunogens and vaccines for
PT stimulating and enhancing immunological responses to target cells and
PT antigens expresses multiple co-stimulatory molecules such as B7-1, LFA-3,
PT ICAM-1.
XX
PS Claim 18; Page 35; 188pp; English.
XX
CC Costimulatory molecules have important roles in T-cell activation and
CC therefore the immune response. The present invention relates to
CC recombinant vectors which comprise of foreign nucleic acid sequences
CC encoding at least three costimulatory molecules: a B7 family molecule,
CC Leukocyte function-associated antigen-3 (LFA-3, human CD58) and
CC Intercellular adhesion molecule-1 (ICAM-1, CD54) and optionally a foreign
CC gene encoding a target antigen or immunological epitope. The present
CC sequence is one such target antigen used in the present invention. The
CC present sequence is a tumour-associated antigen. The vector of the
CC present invention would be useful for providing an enhanced immune
CC response to the present target antigen. The vector of the present
CC invention may therefore be useful in immunotherapy for treating or
CC preventing diseases caused by viruses, bacteria, protozoans, parasites,
CC premalignant cells and tumour cells. The recombinant vector can be used
CC to treat or prevent preneoplastic or hyperplastic states such as colon
CC polyps, Crohn's disease, ulcerative colitis and breast lesions
XX
SQ Sequence 9 AA;

Query Match 75.0%; Score 36; DB 3; Length 9;
Best Local Similarity 88.9%; Pred. No. 1.8e+06;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy          1 YLSGACLNL 9
              ||||| |||
Db          1 YLSGANLNL 9
```

RESULT 11
AAB13750
ID AAB13750 standard; peptide; 9 AA.
XX
AC AAB13750;
XX
DT 02-FEB-2001 (first entry)
XX
DE Peptide fragment # 2 from human CEA.
XX

KW Human; T-cell; immune response; antigen; epitope; B7 family molecule;
KW Leukocyte function-associated antigen-3; LFA-3;
KW Intercellular adhesion molecule-1; ICAM-1; vaccine; immunotherapy;
KW colon polyp; Crohn's disease; ulcerative colitis; breast lesion; tumour;
KW CEA.
XX
OS Homo sapiens.
XX
PN WO200034494-A1.
XX
PD 15-JUN-2000.
XX
PF 12-NOV-1999; 99WO-US026866.
XX
PR 09-DEC-1998; 98US-0111582P.
XX
PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
PA (THER-) THERION BIOLOGICS CORP.
XX
PI Schlom J, Hodge J, Panicali D;
XX
DR WPI; 2000-431307/37.
XX
PT Novel recombinant vector useful as immunogens and vaccines for
PT stimulating and enhancing immunological responses to target cells and
PT antigens expresses multiple co-stimulatory molecules such as B7-1, LFA-3,
PT ICAM-1.
XX
PS Claim 18; Page 35; 188pp; English.
XX
CC Costimulatory molecules have important roles in T-cell activation and
CC therefore the immune response. The present invention relates to
CC recombinant vectors which comprise of foreign nucleic acid sequences
CC encoding at least three costimulatory molecules: a B7 family molecule,
CC Leukocyte function-associated antigen-3 (LFA-3, human CD58) and
CC Intercellular adhesion molecule-1 (ICAM-1, CD54) and optionally a foreign
CC gene encoding a target antigen or immunological epitope. The present
CC sequence is one such target antigen used in the present invention. The
CC present sequence is a tumour-associated antigen. The vector of the
CC present invention would be useful for providing an enhanced immune
CC response to the present target antigen. The vector of the present
CC invention may therefore be useful in immunotherapy for treating or
CC preventing diseases caused by viruses, bacteria, protozoans, parasites,
CC premalignant cells and tumour cells. The recombinant vector can be used
CC to treat or prevent preneoplastic or hyperplastic states such as colon
CC polyps, Crohn's disease, ulcerative colitis and breast lesions
XX
SQ Sequence 9 AA;

Query Match 75.0%; Score 36; DB 3; Length 9;
Best Local Similarity 88.9%; Pred. No. 1.8e+06;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 YLSGACLNL 9
|||||.|||
Db 1 YLSGADLNL 9

RESULT 12
AAE02673
ID AAE02673 standard; peptide; 9 AA.
XX
AC AAE02673;
XX
DT 06-AUG-2001 (first entry)
XX
DE Human CEA epitopic peptide.
XX
KW Human; cytostatic; antibacterial; antifungal; gene therapy; vaccine;
KW antiviral; tumour; epitope; glycoprotein; hepatitis B virus; HBV;
KW immune response; CTL; cytotoxic T lymphocyte; CEA; HLA;
KW human leucocyte antigen.
XX
OS Homo sapiens.
XX
PN WO200127291-A1.
XX
PD 19-APR-2001.
XX
PF 29-SEP-2000; 2000WO-EP009902.
XX
PR 12-OCT-1999; 99US-0158356P.
XX
PA (INSP) INST PASTEUR.
XX
PI Firat H, Lemonnier F, Langlade-Demoyen P;
XX
DR WPI; 2001-282038/29.
XX
PT New polynucleotide comprising at least one viral, fungal, bacterial, or
PT tumor epitope of an antigen, capable of inducing a cellular response.
XX
PS Example 1; Page 23; 70pp; English.
XX
CC The invention relates to polynucleotide containing at least a part of the
CC coding sequence of the middle glycoprotein of hepatitis B virus (HBV) in
CC which is inserted a DNA sequence coding for an epitope comprising at
CC least one viral, fungal, bacterial, or tumour epitope of an antigen,
CC capable of inducing a cellular response. Nucleic acids and compositions
CC of the invention are useful for inducing in vivo a CTL (cytotoxic T
CC lymphocyte) response against several epitopes of one or more, bacterial,
CC viral, fungal, or tumour antigens. A composition of the invention
CC produces an immune response against HIV antigen and are used in the
CC production of vaccines. The polynucleotides of the invention are also
CC used in gene therapy. The present sequence is human CEA epitopic peptide.
CC This peptide elicits strong HLA (human leucocyte antigen)-A2.1-restricted
CC CTL response in mice
XX
SQ Sequence 9 AA;

Query Match 75.0%; Score 36; DB 4; Length 9;
Best Local Similarity 88.9%; Pred. No. 1.8e+06;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 YLSGACLNL 9
|||||.|||
Db 1 YLSGANLNL 9

RESULT 13
AAE00463
ID AAE00463 standard; peptide; 9 AA.
XX
AC AAE00463;
XX
DT 19-JUN-2001 (first entry)
XX
DE Human tumour CEA epitopic peptide.
XX
KW Human; tumour epitope; cytostatic; immunostimulant; gene therapy;
KW middle glycoprotein; Hepatitis B virus; HBV; cytotoxic response;
KW immune response; cytotoxic T lymphocyte; CTL; CEA; HLA;
KW human leucocyte antigen.
XX
OS Homo sapiens.
XX
PN WO200123577-A2.
XX
PD 05-APR-2001.
XX
PF 29-SEP-2000; 2000WO-EP009900.
XX
PR 30-SEP-1999; 99US-0156945P.
XX
PA (INSP) INST PASTEUR.
XX

PI Firat H, Lemonnier F, Langlade-Demoyen P, Michel M, Suhrbier AA;
XX
DR WPI; 2001-266164/27.
XX
PT Novel polynucleotide having DNA sequence encoding tumor antigen epitope
PT inserted in part of coding sequence of middle glycoprotein of hepatitis B
PT virus, used to induce immune response against tumor-specific antigen.
XX
PS Example 1; Page 13; 36pp; English.
XX
CC The present invention relates to an isolated or purified polynucleotide
CC containing a DNA sequence coding for at least one tumour epitope of a
CC tumour antigen inserted into part of the coding sequence of the middle
CC glycoprotein of the Hepatitis B virus (HBV). The polynucleotide is useful
CC for optionally evaluating cytotoxic responses in the individual's
CC lymphocyte population. It induces an immune response against at least one
CC tumour specific antigen or tissue specific antigen. The vector comprising
CC the polynucleotide induces in vivo, cellular and/or humoral immune
CC response. The composition comprising the polynucleotide induces in vivo,
CC cytotoxic T lymphocyte (CTL) against one or more antigens or epitopes
CC present on the hybrid protein. The polynucleotide is also useful in gene
CC therapy. The present sequence is a human tumour CEA epitopic peptide.
CC This peptide elicits strong HLA (human leucocyte antigen)-A2.1-restricted
CC CTL response in mice
XX
SQ Sequence 9 AA;

Query Match 75.0%; Score 36; DB 4; Length 9;
Best Local Similarity 88.9%; Pred. No. 1.8e+06;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 YLSGACLNL 9
||||| |||
Db 1 YLSGANLNL 9

RESULT 14
AAB97818
ID AAB97818 standard; peptide; 9 AA.
XX
AC AAB97818;
XX
DT 08-AUG-2001 (first entry)
XX
DE Carcinoembryonic antigen (CEA) modified antigen SEQ ID NO:113.
XX
KW Virus; adenovirus; poxvirus; alphavirus; immune response; gp100;
KW tumour antigen; CEA; carcinoembryonic antigen; immunostimulant;
KW cytostatic; immunotherapy; interferon-gamma; IFN-gamma; cancer.
XX
OS Unidentified.
XX
PN WO200130382-A1.
XX
PD 03-MAY-2001.
XX
PF 20-OCT-2000; 2000WO-CA001253.
XX
PR 22-OCT-1999; 99US-0160879P.
PR 07-AUG-2000; 2000US-0223325P.
XX
PA (AVET) AVENTIS PASTEUR LTD.
XX
PI Berinstein N, Tartaglia J, Moingeon P, Barber B;
XX
DR WPI; 2001-308587/32.
XX
PT Inducing immune response to tumor antigen, useful in immunotherapy of
PT cancer, by administering the antigen to a lymphatic site.
XX
PS Claim 19; Page 9; 60pp; English.
XX
CC The present invention describes a method for inducing an immune response,
CC in an animal, to a tumour antigen (Ag) comprising administering Ag, or
CC nucleic acid (I) that encodes it, to a lymphatic site. Cynomolgus monkeys
CC (Macaca fascicularis) were injected with a modified form of gp100 antigen
CC (a) into the left inguinal lymph node or (b) subcutaneously. Both animals
CC of (a) developed a cell-mediated response (indicated by production of
CC interferon-gamma from T lymphocytes when exposed to gp100 peptides), but
CC only 2 of 4 animals of (b) did so. Also animals in (a) produced a far
CC greater antibody response to gp100. The method is used in immunotherapy
CC of a wide range of cancers through induction of a specific immune
CC response (humoral and cellular) against the tumour antigens. When
CC administered to a lymphatic site, Ag (or (I)) induces a stronger immune
CC response than administration by other routes and may also break tolerance
CC to Ag. AAB97708 and AAB97709 represent gp100 epitopes; AAB97710 to
CC AAB97815 represent peptides derived from gp100 which stimulate interferon
CC (IFN)-gamma production; AAH20120 encodes the modified gp100 protein given
CC in AAB97816; AAH20121 encodes the modified carcinoembryonic antigen (CEA)
CC protein given in AAB97817; and AAB97818 represents a CEA modified antigen
CC peptide, all of which are used in the exemplification of the present
CC invention
XX
SQ Sequence 9 AA;

Query Match 75.0%; Score 36; DB 4; Length 9;
Best Local Similarity 88.9%; Pred. No. 1.8e+06;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 YLSGACLNL 9
||||| |||
Db 1 YLSGADLNL 9

RESULT 15
AAE05123
ID AAE05123 standard; peptide; 9 AA.
XX
AC AAE05123;
XX
DT 18-SEP-2001 (first entry)
XX
DE Carcinoembryonic antigen (CEA) peptide, CAP-1.
XX
KW Tumour-associated antigen; TAA; cytostatic; vaccine; gene therapy;
KW immune response; tetanus toxoid; TT; diphtheria toxoid; DT; prophylactic;
KW cancer; therapeutic; carcinoembryonic antigen; CEA.
XX
OS Unidentified.
XX
PN WO200149317-A2.
XX
PD 12-JUL-2001.
XX
PF 05-JAN-2001; 2001WO-CA000005.
XX
PR 05-JAN-2000; 2000US-0174587P.
XX
PA (AVET) AVENTIS PASTEUR LTD.
XX
PI Emtage P, Barber BH, Sambhara S, Sia CDY;
XX
DR WPI; 2001-441790/47.
XX
PT Enhancing immune response to antigen such as tumor antigen for treating
PT cancer in an animal involves administering an inducing agent to the
PT animal followed by administering inducing agent-antigen mixture.
XX
PS Example 2; Page 31; 62pp; English.
XX
CC The invention relates to a method of enhancing an immune response against
CC tumour-associated antigens (TAAs), such as GP100 and carcinoembryonic
CC antigen (CEA) in an animal. The method involves priming of the animal
CC with an inducing agent such as tetanus toxoid (TT) or diphtheria toxoid
CC (DT), subsequently followed by administration of an inducing agent-
CC antigen mixture. The method provides the enhancement or augmentation of

CC the immune response to the antigen and/or improves a vaccination protocol
CC by allowing use of less antigen. The immunisation of the animal with
CC tumour-associated antigen is useful for the prophylactic or therapeutic
CC treatment of cancer. The present sequence is carcinoembryonic antigen
CC (CEA) peptide fragment related to the invention
XX
SQ Sequence 9 AA;

Query Match 75.0%; Score 36; DB 4; Length 9;
Best Local Similarity 88.9%; Pred. No. 1.8e+06;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 YLSGACLNL 9
||||| |||
Db 1 YLSGANLNL 9

Search completed: May 17, 2005, 06:17:51
Job time : 66 secs

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: May 17, 2005, 06:13:14 ; Search time 11.25 Seconds
(without alignments)
76.973 Million cell updates/sec

Title: US-10-725-373-3
Perfect score: 45
Sequence: 1 YLSGADINL 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_79:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	38	84.4	702	2	A36319	carcinoembryonic a
2	34	75.6	176	2	F71064	micrococcal nuclea
3	34	75.6	275	2	PN0511	gastrin-binding pr
4	34	75.6	714	2	C65007	probable fatty oxi
5	34	75.6	714	2	A85876	probable enzyme Z3
6	34	75.6	714	2	H91031	probable enzyme [i
7	34	75.6	715	2	AB0805	3-hydroxyacyl-CoA
8	34	75.6	763	2	JC2108	long-chain-fatty-a
9	34	75.6	834	2	F82673	hypothetical prote
10	34	75.6	860	2	C72338	hypothetical prote
11	34	75.6	953	2	AH1972	hypothetical prote
12	34	75.6	1986	2	S28353	probable polyketid
13	33	73.3	176	2	AH0881	probable exported
14	33	73.3	311	2	E43680	D'311 protein - Af
15	33	73.3	316	2	C86828	2-dehydro-3-deoxyg
16	33	73.3	401	2	T24929	hypothetical prote
17	33	73.3	402	2	T24930	hypothetical prote
18	33	73.3	441	2	T20840	hypothetical prote
19	33	73.3	445	2	T21262	hypothetical prote
20	33	73.3	454	2	T21261	hypothetical prote
21	33	73.3	705	2	T12784	sublancin 168 lant
22	33	73.3	763	1	A49681	long-chain-fatty-a
23	33	73.3	880	2	S51473	probable membrane
24	33	73.3	1006	2	D86431	protein T5I8.6 [im
25	32	71.1	188	2	G70475	conserved hypothet
26	32	71.1	281	2	E90112	26S proteasome reg
27	32	71.1	312	2	T33344	hypothetical prote
28	32	71.1	350	2	T22450	hypothetical prote
29	32	71.1	363	2	JC4748	polygalacturonase
30	32	71.1	374	2	E69049	hypothetical prote
31	32	71.1	391	2	A99111	26S proteasome AAA
32	32	71.1	414	2	A88485	protein F23F12.6 [
33	32	71.1	506	2	C81704	monooxygenase-rela
34	32	71.1	540	2	E69861	ABC transporter (A
35	32	71.1	816	2	AI2444	hypothetical prote
36	32	71.1	1023	2	T30257	IgG Fc binding pro
37	32	71.1	1650	2	S53457	dominant autoantig
38	32	71.1	4660	2	T42737	gp330 protein prec
39	31	68.9	30	2	A03148	retinol-binding pr
40	31	68.9	108	2	S37313	transcription acti
41	31	68.9	112	2	PQ0493	hypothetical prote
42	31	68.9	172	1	A41841	phycoerythrocyanin
43	31	68.9	172	2	AB1872	phycoerythrocyanin
44	31	68.9	175	2	D75083	micrococcal nuclea
45	31	68.9	210	2	S08389	hypothetical prote

ALIGNMENTS

RESULT 1
A36319
carcinoembryonic antigen precursor - human
N;Alternate names: CEA; meconium antigen 100
C;Species: Homo sapiens (man)
C;Date: 16-Sep-1992 #sequence_revision 16-Sep-1992 #text_change 09-Jul-2004
C;Accession: A36319; A27773; A31037; A25845; S08106; S31737; A44476; I54224; I59098; A26
R;Schrewe, H.; Thompson, J.; Bona, M.; Hefta, L.J.F.; Maruya, A.; Hassauer, M.; Shively,
Mol. Cell. Biol. 10, 2738-2748, 1990
A;Title: Cloning of the complete gene for carcinoembryonic antigen: analysis of its promo
A;Reference number: A36319; MUID:90258861; PMID:2342461
A;Accession: A36319
A;Molecule type: DNA
A;Residues: 1-702 <SCH>
A;Cross-references: UNIPROT:P06731; GB:M17303; NID:g178676; PIDN:AAB59513.1; PID:g178677
A;Note: the authors show the codons TTA for residue 641-Phe and CAG for residue 646-Thr
R;Beauchemin, N.; Benchimol, S.; Cournoyer, D.; Fuks, A.; Stanners, C.P.
Mol. Cell. Biol. 7, 3221-3230, 1987
A;Title: Isolation and characterization of full-length functional cDNA clones for human c
A;Reference number: A27773; MUID:88038876; PMID:3670312
A;Accession: A27773
A;Molecule type: mRNA
A;Residues: 1-702 <BEA>
A;Cross-references: GB:M29540; NID:g180222; PIDN:AAA51967.1; PID:g180223
R;Barnett, T.; Goebel, S.J.; Nothdurft, M.A.; Elting, J.J.
Genomics 3, 59-66, 1988
A;Title: Carcinoembryonic antigen family: characterization of cDNAs coding for NCA and CE
A;Reference number: A31037; MUID:89122014; PMID:3220478
A;Accession: A31037
A;Molecule type: mRNA
A;Residues: 1-702 <BAR>
A;Cross-references: GB:M29540; NID:g180222; PIDN:AAA51967.1; PID:g180223
A;Note: the authors translated the codon GTG for residue 130 as Leu
R;Oikawa, S.; Nakazato, H.; Kosaki, G.
Biochem. Biophys. Res. Commun. 142, 511-518, 1987
A;Title: Primary structure of human carcinoembryonic antigen (CEA) deduced from cDNA sequ
A;Reference number: A25845; MUID:87128144; PMID:3814146
A;Accession: A25845
A;Molecule type: mRNA
A;Residues: 5-702 <OIK>
A;Cross-references: GB:M15042; NID:g180198; PIDN:AAA51963.1; PID:g180199
R;Oikawa, S.
submitted to the EMBL Data Library, September 1989
A;Reference number: S08106
A;Accession: S08106
A;Molecule type: mRNA
A;Residues: 5-319,321-702 <OI2>
A;Cross-references: EMBL:X16455; NID:g29854; PIDN:CAA34474.1; PID:g825638
R;Barnett, T.
submitted to the EMBL Data Library, September 1991
A;Description: Genomic DNA sequence upstream of the translational start of the carcinoemb
A;Reference number: S31737

A;Accession: S31737
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-141 <BA2>
A;Cross-references: EMBL:X62151
R;Khan, W.N.; Fraengsmyr, L.; Teglund, S.; Israelsson, A.; Bremer, K.; Hammarstroem, S.
Genomics 14, 384-390, 1992
A;Title: Identification of three new genes and estimation of the size of the carcinoembr
A;Reference number: A44476; MUID:93052339; PMID:1427854
A;Accession: A44476
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: DNA
A;Residues: 35-141 <KHA>
R;Willcocks, T.C.; Craig, I.W.
Genomics 8, 492-500, 1990
A;Title: Characterization of the genomic organization of human carcinoembryonic antigen
A;Reference number: I54224; MUID:91139118; PMID:2286372
A;Accession: I54224
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-37 <RES>
A;Cross-references: GB:M60964; NID:g180215; PIDN:AAA51964.1; PID:g180217
R;Zimmermann, W.; Ortlieb, B.; Friedrich, R.; von Kleist, S.
Proc. Natl. Acad. Sci. U.S.A. 84, 2960-2964, 1987
A;Title: Isolation and characterization of cDNA clones encoding the human carcinoembryon
A;Reference number: I59098; MUID:87204247; PMID:3033671
A;Accession: I59098
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 331-702 <RE2>
A;Cross-references: GB:M16234; NID:g180240; PIDN:AAA51972.1; PID:g180241
R;Siepen, D.; Paxton, R.J.; Neumaier, M.; Shively, J.E.; Wagener, C.
Biochem. Biophys. Res. Commun. 147, 212-218, 1987
A;Title: Carcinoembryonic antigen (CEA) and two crossreacting antigens of 165 KD and 105
A;Reference number: A26831; MUID:87326349; PMID:3632664
A;Accession: A26831
A;Molecule type: protein
A;Residues: 35-64 <SIE>
R;Thomas, P.; Toth, C.A.
Biochem. Biophys. Res. Commun. 170, 391-396, 1990
A;Title: Carcinoembryonic antigen binding to Kupffer cells is via a peptide located at t
A;Reference number: A35490; MUID:90321257; PMID:2372297
A;Accession: A35490
A;Molecule type: protein
A;Residues: 'X',140-151,'X',153,'X',155-156 <THO>
A;Note: this is the amino terminal end of a fragment shown to mediate uptake by Kupffer
C;Comment: This heavily glycosylated membane protein of unknown function is a widely use
C;Comment: This protein may be processed at its C-terminus. It is anchored to the membra
C;Genetics:
A;Gene: GDB:CEA
A;Cross-references: GDB:119054; OMIM:114890
A;Map position: 19q13.2-19q13.2
A;Introns: 22/1; 142/1; 235/1; 320/1; 413/1; 498/1; 591/1; 676/1
C;Superfamily: carcinoembryonic antigen; carcinoembryonic antigen precursor amino-termin
C;Keywords: blocked carboxyl end; glycoprotein; lipoprotein; membrane protein; phosphati
F;1-138/Domain: carcinoembryonic antigen precursor amino-terminal homology <CEAN>
F;1-34/Domain: signal sequence #status predicted <SIG>
F;35-678/Product: carcinoembryonic antigen #status predicted <MAT>
F;160-217/Domain: immunoglobulin homology <IMM1>
F;252-301/Domain: immunoglobulin homology <IMM2>
F;338-395/Domain: immunoglobulin homology <IMM3>
F;516-573/Domain: immunoglobulin homology <IMM4>
F;608-657/Domain: immunoglobulin homology <IMM5>
F;679-702/Domain: carboxyl-terminal propeptide #status predicted <CTP>
F;678/Modified site: GPI-anchor ethanolamine amidated carboxyl end (Gly) (in mature form

  Query Match 84.4%; Score 38; DB 2; Length 702;
  Best Local Similarity 77.8%; Pred. No. 14;
  Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 YLSGADINL 9
||||||::||
Db 605 YLSGANLNL 613

RESULT 2
F71064
micrococcal nuclease (EC 3.1.31.1) PH1212 precursor - Pyrococcus horikoshii
N;Alternate names: thermonuclease homolog
C;Species: Pyrococcus horikoshii
C;Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 12-Jul-2004
C;Accession: F71064
R;Kawarabayasi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; Sekine
M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguchi,
DNA Res. 5, 55-76, 1998
A;Title: Complete sequence and gene organization of the genome of a hyper-thermophilic ar
A;Reference number: A71000; MUID:98344137; PMID:9679194
A;Accession: F71064
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-176 <KAW>
A;Cross-references: UNIPROT:O58971; GB:AP000005; NID:g3236132; PIDN:BAA30312.1; PID:g3257
A;Experimental source: strain OT3
A;Note: this accession replaces an interim accession for a sequence replaced by GenBank
C;Genetics:
A;Gene: PH1212
C;Keywords: hydrolase
F;1-27/Domain: signal sequence #status predicted <SIG>

  Query Match 75.6%; Score 34; DB 2; Length 176;
  Best Local Similarity 75.0%; Pred. No. 20;
  Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 YLSGADIN 8
||:| |||
Db 128 YLNGTDIN 135

RESULT 3
PN0511
gastrin-binding protein precursor - pig (fragment)
C;Species: Sus scrofa domestica (domestic pig)
C;Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 17-Mar-1999
C;Accession: PN0511
R;Baldwin, G.S.; Casey, A.; Weinstock, J.
Biochem. Biophys. Res. Commun. 193, 560-564, 1993
A;Title: Partial structure of the gene encoding the 78KDa gastrin binding protein enclude
A;Reference number: PN0511; MUID:93290643; PMID:8512557
A;Accession: PN0511
A;Molecule type: mRNA
A;Residues: 1-275 <BAL>
A;Note: complete nucleotide sequence not given
C;Genetics:
A;Introns: 23/1; 37/1; 60/3; 105/2; 151/3; 191/3; 226/1
C;Superfamily: enoyl-CoA hydratase/3-hydroxyacyl-CoA dehydrogenase; 3-hydroxyacyl-CoA deh
F;1-36/Domain: signal sequence #status predicted <SIG>
F;37-275/Product: gastrin-binding protein #status predicted <MAT>
F;62-218/Domain: enoyl-CoA hydratase homology <ECH>

  Query Match 75.6%; Score 34; DB 2; Length 275;
  Best Local Similarity 55.6%; Pred. No. 33;
  Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 1 YLSGADINL 9
:::|||||:
Db 98 FIAGADINM 106

RESULT 4
C65007
probable fatty oxidation complex alpha subunit - Escherichia coli (strain K-12)
C;Species: Escherichia coli
C;Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 09-Jul-2004
C;Accession: C65007
R;Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Col
.A.; Rose, D.J.; Mau, B.; Shao, Y.

Science 277, 1453-1462, 1997
A;Title: The complete genome sequence of Escherichia coli K-12.
A;Reference number: A64720; MUID:97426617; PMID:9278503
A;Accession: C65007
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-714 <BLAT>
A;Cross-references: UNIPROT:P77399; GB:AE000322; GB:U00096; NID:g1788672; PIDN:AAC75401.
A;Experimental source: strain K-12, substrain MG1655
C;Superfamily: enoyl-CoA hydratase/3-hydroxyacyl-CoA dehydrogenase; 3-hydroxyacyl-CoA de
F;29-185/Domain: enoyl-CoA hydratase homology <ECH>
F;308-587/Domain: 3-hydroxyacyl-CoA dehydrogenase homology <HCD>

Query Match 75.6%; Score 34; DB 2; Length 714;
Best Local Similarity 55.6%; Pred. No. 91;
Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 1 YLSGADINL 9
:::|||||:
Db 65 FIAGADINM 73

RESULT 5
A85876
probable enzyme Z3604 [imported] - Escherichia coli (strain O157:H7, substrain EDL933)
C;Species: Escherichia coli
C;Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 09-Jul-2004
C;Accession: A85876
R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
Nature 409, 529-533, 2001
A;Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A;Reference number: A85480; MUID:21074935; PMID:11206551
A;Accession: A85876
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-714 <STO>
A;Cross-references: UNIPROT:Q8XCP2; GB:AE005174; NID:g12516705; PIDN:AAG57469.1; GSPDB:G
A;Experimental source: strain O157:H7, substrain EDL933
C;Genetics:
A;Gene: Z3604
C;Superfamily: enoyl-CoA hydratase/3-hydroxyacyl-CoA dehydrogenase; 3-hydroxyacyl-CoA de

Query Match 75.6%; Score 34; DB 2; Length 714;
Best Local Similarity 55.6%; Pred. No. 91;
Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 1 YLSGADINL 9
:::|||||:
Db 65 FIAGADINM 73

RESULT 6
H91031
probable enzyme [imported] - Escherichia coli (strain O157:H7, substrain RIMD 0509952)
C;Species: Escherichia coli
C;Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 09-Jul-2004
C;Accession: H91031
R;Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A;Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genc
A;Reference number: A99629; MUID:21156231; PMID:11258796
A;Accession: H91031
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-714 <HAY>
A;Cross-references: UNIPROT:Q8XCP2; GB:BA000007; PIDN:BAB36647.1; PID:g13362694; GSPDB:G
A;Experimental source: strain O157:H7, substrain RIMD 0509952
C;Genetics:
A;Gene: ECs3224
C;Superfamily: enoyl-CoA hydratase/3-hydroxyacyl-CoA dehydrogenase; 3-hydroxyacyl-CoA de

Query Match 75.6%; Score 34; DB 2; Length 714;
Best Local Similarity 55.6%; Pred. No. 91;
Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 1 YLSGADINL 9
:::|||||:
Db 65 FIAGADINM 73

RESULT 7
AB0805
3-hydroxyacyl-CoA dehydrogenase (EC 1.1.1.35) - Salmonella enterica subsp. enterica serov
C;Species: Salmonella enterica subsp. enterica serovar Typhi
A;Note: this species has also been called Salmonella typhi
C;Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002
C;Accession: AB0805
R;Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,
th, T.; Connerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,
, S.; Moule, S.; O'Gaora, P.
Nature 413, 848-852, 2001
A;Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;
A;Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serova
A;Reference number: AB0502; MUID:21534947; PMID:11677608
A;Accession: AB0805
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-715 <PAR>
A;Cross-references: GB:AL513382; PIDN:CAD07620.1; PID:g16503611; GSPDB:GN00176
C;Genetics:
A;Gene: STY2620
C;Superfamily: enoyl-CoA hydratase/3-hydroxyacyl-CoA dehydrogenase; 3-hydroxyacyl-CoA deh
C;Keywords: oxidoreductase

Query Match 75.6%; Score 34; DB 2; Length 715;
Best Local Similarity 55.6%; Pred. No. 91;
Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 1 YLSGADINL 9
:::|||||:
Db 65 FIAGADINM 73

RESULT 8
JC2108
long-chain-fatty-acid beta-oxidation multienzyme complex alpha chain precursor, mitochonc
N;Alternate names: 78K gastrin-binding protein
N;Contains: long-chain-3-hydroxyacyl-CoA dehydrogenase (EC 1.1.1.211); long-chain-enoyl-C
C;Species: Homo sapiens (man)
C;Date: 14-Jul-1994 #sequence_revision 14-Jul-1994 #text_change 09-Jul-2004
C;Accession: JC2108; PC2058; S50127
R;Kamijo, T.; Aoyama, T.; Komiyama, A.; Hashimoto, T.
Biochem. Biophys. Res. Commun. 199, 818-825, 1994
A;Title: Structural analysis of cDNAs for subunits of human mitochondrial fatty acid beta
A;Reference number: JC2108; MUID:94183263; PMID:8135828
A;Accession: JC2108
A;Molecule type: mRNA
A;Residues: 1-763 <KAM>
A;Cross-references: UNIPROT:P40939; GB:D16480; NID:g493657; PIDN:BAA03941.1; PID:g862457
A;Accession: PC2058
A;Molecule type: protein
A;Residues: 37-64 <KA2>
A;Note: peptide sequence from amino end of mature protein
R;Zhang, Q.X.; Baldwin, G.S.
Biochim. Biophys. Acta 1219, 567-575, 1994
A;Title: Structures of the human cDNA and gene encoding the 78 kDa gastrin-binding protei
A;Reference number: S50127; MUID:95002180; PMID:7918661
A;Accession: S50127
A;Status: preliminary; translation not shown
A;Molecule type: mRNA
A;Residues: 1-145,'L',147-151,'L',153-170,'A',172-177,'I',179-196,'VF',199-205,'N',207-21
A;Cross-references: EMBL:U04627; NID:g595266; PIDN:AAA56664.1; PID:g595267
C;Genetics:
A;Gene: GDB:HADHA

A;Cross-references: GDB:434026; OMIM:600890
A;Map position: 2p23-2p23
C;Complex: heterooctamer of 4 alpha and 4 beta chains
C;Superfamily: enoyl-CoA hydratase/3-hydroxyacyl-CoA dehydrogenase; 3-hydroxyacyl-CoA de
C;Keywords: acyltransferase; carbon-oxygen lyase; fatty acid beta-oxidation; heterooctam
F;1-36/Domain: transit peptide (mitochondrion) #status predicted <TPP>
F;37-763/Product: fatty acid beta-oxidation trifunctional protein, alpha chain #status e
F;62-218/Domain: enoyl-CoA hydratase homology <ECH>
F;361-640/Domain: 3-hydroxyacyl-CoA dehydrogenase homology <HCD>
F;363-391/Region: beta-alpha-beta NAD nucleotide-binding fold

Query Match 75.6%; Score 34; DB 2; Length 763;
Best Local Similarity 55.6%; Pred. No. 98;
Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 1 YLSGADINL 9
:::|||||:
Db 98 FIAGADINM 106

RESULT 9
F82673
hypothetical protein XF1508 [imported] - Xylella fastidiosa (strain 9a5c)
C;Species: Xylella fastidiosa
C;Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 09-Jul-2004
C;Accession: F82673
R;anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequen
Nature 406, 151-157, 2000
A;Title: The genome sequence of the plant pathogen Xylella fastidiosa.
A;Reference number: A82515; MUID:20365717; PMID:10910347
A;Note: for a complete list of authors see reference number A59328 below
A;Accession: F82673
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-834 <SIM>
A;Cross-references: UNIPROT:Q9PD71; GB:AE003980; GB:AE003849; NID:g9106531; PIDN:AAF8431
A;Experimental source: strain 9a5c
R;Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A
Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrer, H
as-Neto, E.; Docena, C.; El-Dorry, H.; Facincani, A.P.; Ferreira, A.J.S.
submitted to GenBank, June 2000
A;Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm
J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laigr
chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E
A;Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;
, F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak
A;Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir
M.; Tsuhako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z
A;Reference number: A59328
A;Contents: annotation
C;Genetics:
A;Gene: XF1508

Query Match 75.6%; Score 34; DB 2; Length 834;
Best Local Similarity 75.0%; Pred. No. 1.1e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 YLSGADIN 8
|||| |:|
Db 391 YLSGMDLN 398

RESULT 10
C72338
hypothetical protein - Thermotoga maritima (strain MSB8)
C;Species: Thermotoga maritima
C;Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 09-Jul-2004
C;Accession: C72338
R;Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hickey
Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.;
C.M.
Nature 399, 323-329, 1999
A;Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome sequ
A;Reference number: A72200; MUID:99287316; PMID:10360571
A;Accession: C72338
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-860 <ARN>
A;Cross-references: UNIPROT:Q9WZL6; GB:AE001745; GB:AE000512; NID:g4981278; PIDN:AAD3583
A;Experimental source: strain MSB8
C;Genetics:
A;Gene: TM0757

Query Match 75.6%; Score 34; DB 2; Length 860;
Best Local Similarity 75.0%; Pred. No. 1.1e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 YLSGADIN 8
||:| |||
Db 443 YLTGGDIN 450

RESULT 11
AH1972
hypothetical protein alr1331 [imported] - Nostoc sp. (strain PCC 7120)
C;Species: Nostoc sp. PCC 7120
A;Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C;Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Jul-2004
C;Accession: AH1972
R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi,
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.
DNA Res. 8, 205-213, 2001
A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anab
A;Reference number: AB1807; MUID:21595285; PMID:11759840
A;Accession: AH1972
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-953 <KUR>
A;Cross-references: UNIPROT:Q8YX84; GB:BA000019; PIDN:BAB73288.1; PID:g17130678; GSPDB:GN
A;Experimental source: strain PCC 7120
C;Genetics:
A;Gene: alr1331

Query Match 75.6%; Score 34; DB 2; Length 953;
Best Local Similarity 75.0%; Pred. No. 1.2e+02;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 YLSGADIN 8
||||||::
Db 828 YLSGADLS 835

RESULT 12
S28353
probable polyketide synthase - Emericella nidulans
C;Species: Emericella nidulans, Aspergillus nidulans
C;Date: 17-Apr-1993 #sequence_revision 17-Apr-1993 #text_change 09-Jul-2004
C;Accession: S28353
R;Mayorga, M.E.; Timberlake, W.E.
Mol. Gen. Genet. 235, 205-212, 1992
A;Title: The developmentally regulated Aspergillus nidulans wA gene encodes a polypeptide
A;Reference number: S28353; MUID:93101122; PMID:1465094
A;Accession: S28353
A;Molecule type: DNA
A;Residues: 1-1986 <MAY>
A;Cross-references: UNIPROT:Q03149; EMBL:X65866; NID:g5508; PID:g5509
C;Genetics:
A;Gene: wA
A;Introns: 96/2; 193/3; 1336/3; 1588/3
C;Keywords: carrier protein
F;397-805/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS>
F;911-1199/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT>
F;1648-1718/Domain: acyl carrier protein homology <ACP>
F;1766-1840/Domain: acyl carrier protein homology <ACP1>

```
  Query Match             75.6%;  Score 34;  DB 2;  Length 1986;
  Best Local Similarity   75.0%;  Pred. No. 2.7e+02;
  Matches    6;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 YLSGADIN 8
              ||:| |||
Db       1227 YLAGVDIN 1234


RESULT 13
AH0881
probable exported protein STY3288 [imported] - Salmonella enterica subsp. enterica serov
C;Species: Salmonella enterica subsp. enterica serovar Typhi
A;Note: this species has also been called Salmonella typhi
C;Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002
C;Accession: AH0881
R;Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,
th, T.; Connerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,
, S.; Moule, S.; O'Gaora, P.
Nature 413, 848-852, 2001
A;Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;
A;Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov
A;Reference number: AB0502; MUID:21534947; PMID:11677608
A;Accession: AH0881
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-176 <PAR>
A;Cross-references: GB:AL513382; PIDN:CAD02952.1; PID:g16504200; GSPDB:GN00176
C;Genetics:
A;Gene: STY3288

  Query Match             73.3%;  Score 33;  DB 2;  Length 176;
  Best Local Similarity   66.7%;  Pred. No. 33;
  Matches    6;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 YLSGADINL 9
              | || |||:
Db         41 YYSGKDINI 49


RESULT 14
E43680
D'311 protein - African swine fever virus (strain BA71V)
C;Species: African swine fever virus, ASFV
C;Date: 28-Apr-1993 #sequence_revision 28-Apr-1993 #text_change 09-Jul-2004
C;Accession: E43680
R;Gonzalez, A.; Calvo, V.; Almazan, F.; Almendral, J.M.; Ramirez, J.C.; De La Vega, I.;
J. Virol. 64, 2073-2081, 1990
A;Title: Multigene families in African swine fever virus: family 360.
A;Reference number: A43680; MUID:90219205; PMID:2325203
A;Accession: E43680
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-311 <GON>
A;Cross-references: UNIPROT:P23163; GB:M57546
C;Superfamily: African swine fever virus L356 protein

  Query Match             73.3%;  Score 33;  DB 2;  Length 311;
  Best Local Similarity   75.0%;  Pred. No. 60;
  Matches    6;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 YLSGADIN 8
              |: |||||
Db        258 YILGADIN 265


RESULT 15
C86828
2-dehydro-3-deoxygluconokinase (EC 2.7.1.45) [imported] - Lactococcus lactis subsp. lact
C;Species: Lactococcus lactis subsp. lactis
C;Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 09-Jul-2004
C;Accession: C86828
R;Bolotin, A.; Wincker, P.; Mauger, S.; Jaillon, O.; Malarme, K.; Weissenbach, J.; Ehrli
Genome Res. 11, 731-753, 2001
A;Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis ssp
A;Reference number: A86625; MUID:21235186; PMID:11337471
A;Accession: C86828
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-316 <STO>
A;Cross-references: UNIPROT:Q9CF54; GB:AE005176; PID:g12724636; PIDN:AAK05725.1; GSPDB:GN
A;Experimental source: strain IL1403
C;Genetics:
A;Gene: kdgK
C;Superfamily: ribokinase
C;Keywords: phosphotransferase

  Query Match             73.3%;  Score 33;  DB 2;  Length 316;
  Best Local Similarity   55.6%;  Pred. No. 61;
  Matches    5;  Conservative    4;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 YLSGADINL 9
              ||:||::|:
Db         32 YLAGAELNV 40


Search completed: May 17, 2005, 06:20:02
Job time : 13.25 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: May 17, 2005, 06:13:14 ; Search time 51.75 Seconds
(without alignments)
89.057 Million cell updates/sec

Title: US-10-725-373-3
Perfect score: 45
Sequence: 1 YLSGADINL 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : UniProt_03:*
1: uniprot_sprot:*
2: uniprot_trembl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	38	84.4	420	2	Q68DM9	Q68dm9 homo sapien
2	38	84.4	702	1	CEA5_HUMAN	P06731 homo sapien
3	38	84.4	702	2	Q8N4D0	Q8n4d0 homo sapien
4	36	80.0	455	1	PEX3_PICPA	Q92262 pichia past
5	35	77.8	322	2	Q8A067	Q8a067 bacteroides
6	35	77.8	352	2	Q65258	Q65258 african swi
7	35	77.8	388	2	Q948C5	Q948c5 oryza sativ
8	35	77.8	388	2	Q7XGY0	Q7xgy0 oryza sativ
9	35	77.8	499	2	Q6PW77	Q6pw77 acremonium
10	35	77.8	856	2	Q7NMP7	Q7nmp7 gloeobacter
11	34	75.6	131	2	Q6ZWE8	Q6zwe8 homo sapien
12	34	75.6	152	2	Q857N2	Q857n2 mycobacteri
13	34	75.6	176	2	O58971	O58971 pyrococcus
14	34	75.6	237	2	Q7U4R6	Q7u4r6 synechococc
15	34	75.6	268	2	Q96LU9	Q96lu9 homo sapien
16	34	75.6	284	2	Q7NL68	Q7nl68 gloeobacter
17	34	75.6	301	2	Q8N6D5	Q8n6d5 homo sapien
18	34	75.6	317	2	Q9BX27	Q9bx27 homo sapien
19	34	75.6	323	2	Q7RLP1	Q7rlp1 plasmodium
20	34	75.6	336	1	SYFA_TROW8	Q83hh3 tropheryma
21	34	75.6	336	1	SYFA_TROWT	Q83gs9 tropheryma
22	34	75.6	498	2	Q8BNB0	Q8bnb0 mus musculu
23	34	75.6	572	1	GPC5_HUMAN	P78333 homo sapien
24	34	75.6	572	2	Q8CAL5	Q8cal5 mus musculu
25	34	75.6	703	2	Q8NAN1	Q8nan1 homo sapien
26	34	75.6	709	2	Q96GS8	Q96gs8 homo sapien
27	34	75.6	709	2	Q9NUQ8	Q9nuq8 homo sapien
28	34	75.6	709	2	Q8K268	Q8k268 mus musculu
29	34	75.6	709	2	Q66H39	Q66h39 rattus norv
30	34	75.6	711	2	Q86UA2	Q86ua2 homo sapien
31	34	75.6	714	1	YFCX_ECOLI	P77399 escherichia
32	34	75.6	714	2	Q6KCX2	Q6kcx2 escherichia
33	34	75.6	714	2	Q8FFG4	Q8ffg4 escherichia
34	34	75.6	714	2	Q8XCP2	Q8xcp2 escherichia
35	34	75.6	714	2	Q83QQ0	Q83qq0 shigella fl
36	34	75.6	715	2	Q8Z4Z0	Q8z4z0 salmonella
37	34	75.6	715	2	Q8ZNA7	Q8zna7 salmonella
38	34	75.6	760	2	Q6GPS9	Q6gps9 xenopus lae
39	34	75.6	760	2	Q6P4Y3	Q6p4y3 xenopus tro
40	34	75.6	763	1	ECHA_HUMAN	P40939 homo sapien
41	34	75.6	763	1	ECHA_PIG	Q29554 sus scrofa
42	34	75.6	817	2	Q752X5	Q752x5 ashbya goss
43	34	75.6	828	2	Q87DG0	Q87dg0 xylella fas
44	34	75.6	834	2	Q9PD71	Q9pd71 xylella fas
45	34	75.6	846	2	Q9AIP5	Q9aip5 carsonella

ALIGNMENTS

RESULT 1
Q68DM9
ID Q68DM9 PRELIMINARY; PRT; 420 AA.
AC Q68DM9;
DT 25-OCT-2004 (TrEMBLrel. 28, Created)
DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE Hypothetical protein DKFZp781M2392.
GN Name=DKFZp781M2392;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Colon carcinoma;
RG The German cDNA Consortium;
RA Poustka A., Albert R., Moosmayer P., Schupp I., Wellenreuther R.,
RA Mewes H.W., Weil B., Amid C., Osanger A., Fobo G., Han M., Wiemann S.;
RL Submitted (AUG-2004) to the EMBL/GenBank/DDBJ databases.
DR EMBL; CR749337; CAH18191.1; -.
DR InterPro; IPR001589; Actbind_actnin.
DR InterPro; IPR003599; Ig.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003598; Ig_c2.
DR Pfam; PF00047; ig; 3.
DR SMART; SM00409; IG; 3.
DR SMART; SM00408; IGc2; 3.
DR PROSITE; PS00019; ACTININ_1; UNKNOWN_1.
DR PROSITE; PS50835; IG_LIKE; 3.
KW Hypothetical protein.
SQ SEQUENCE 420 AA; 45508 MW; 6E30C0B4A00D0F59 CRC64;

Query Match 84.4%; Score 38; DB 2; Length 420;
Best Local Similarity 77.8%; Pred. No. 49;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 YLSGADINL 9
              |||||::||
Db        323 YLSGANLNL 331
```

RESULT 2
CEA5_HUMAN
ID CEA5_HUMAN STANDARD; PRT; 702 AA.
AC P06731;
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 25-OCT-2004 (Rel. 45, Last annotation update)
DE Carcinoembryonic antigen-related cell adhesion molecule 5 precursor
DE (Carcinoembryonic antigen) (CEA) (Meconium antigen 100) (CD66e
DE antigen).
GN Name=CEACAM5; Synonyms=CEA;
OS Homo sapiens (Human).

```
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=90258861; PubMed=2342461;
RA   Schrewe H., Thompson J., Bona M., Hefta L.J.F., Maruya A.,
RA   Hassauer M., Shively J.E., von Kleist S., Zimmermann W.;
RT   "Cloning of the complete gene for carcinoembryonic antigen: analysis
RT   of its promoter indicates a region conveying cell type-specific
RT   expression.";
RL   Mol. Cell. Biol. 10:2738-2748(1990).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=88038876; PubMed=3670312;
RA   Beauchemin N., Benchimol S., Cournoyer D., Fuks A., Stanners C.P.;
RT   "Isolation and characterization of full-length functional cDNA clones
RT   for human carcinoembryonic antigen.";
RL   Mol. Cell. Biol. 7:3221-3230(1987).
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=89122014; PubMed=3220478;
RA   Barnett T., Goebel S.J., Nothdurft M.A., Elting J.J.;
RT   "Carcinoembryonic antigen family: characterization of cDNAs coding for
RT   NCA and CEA and suggestion of nonrandom sequence variation in their
RT   conserved loop-domains.";
RL   Genomics 3:59-66(1988).
RN   [4]
RP   SEQUENCE OF 5-702 FROM N.A.
RX   MEDLINE=87128144; PubMed=3814146;
RA   Oikawa S., Nakazato H., Kosaki G.;
RT   "Primary structure of human carcinoembryonic antigen (CEA) deduced
RT   from cDNA sequence.";
RL   Biochem. Biophys. Res. Commun. 142:511-518(1987).
RN   [5]
RP   SEQUENCE OF 331-702 FROM N.A.
RX   MEDLINE=87204247; PubMed=3033671;
RA   Zimmermann W., Ortlieb B., Friedrich R., von Kleist S.;
RT   "Isolation and characterization of cDNA clones encoding the human
RT   carcinoembryonic antigen reveal a highly conserved repeating
RT   structure.";
RL   Proc. Natl. Acad. Sci. U.S.A. 84:2960-2964(1987).
CC   -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.
CC   -!- TISSUE SPECIFICITY: Found in adenocarcinomas of endodermally
CC       derived digestive system epithelium and fetal colon.
CC   -!- PTM: Complex immunoreactive glycoprotein with a MW of 180 kDa
CC       comprising 60% carbohydrate.
CC   -!- SIMILARITY: Belongs to the immunoglobulin superfamily. CEA family.
CC   -!- SIMILARITY: Contains 7 immunoglobulin-like domains.
CC   -!- DATABASE: NAME=PROW; NOTE=CD guide CD66e entry;
CC       WWW="http://www.ncbi.nlm.nih.gov/prow/cd/cd66e.htm".
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M17303; AAB59513.1; -.
DR   EMBL; M59262; AAA62835.1; ALT_SEQ.
DR   EMBL; M59255; AAA62835.1; JOINED.
DR   EMBL; M59256; AAA62835.1; JOINED.
DR   EMBL; M59257; AAA62835.1; JOINED.
DR   EMBL; M59258; AAA62835.1; JOINED.
DR   EMBL; M59259; AAA62835.1; JOINED.
DR   EMBL; M59260; AAA62835.1; JOINED.
DR   EMBL; M59261; AAA62835.1; JOINED.
DR   EMBL; M59709; -; NOT_ANNOTATED_CDS.
DR   EMBL; M59710; -; NOT_ANNOTATED_CDS.
DR   EMBL; M29540; AAA51967.1; -.
DR   EMBL; X16455; CAA34474.1; -.
DR   EMBL; M15042; AAA51963.1; -.
DR   EMBL; M16234; AAA51972.1; -.
DR   PIR; A36319; A36319.
DR   PDB; 1E07; Model; A=35-676.
DR   Genew; HGNC:1817; CEACAM5.
DR   MIM; 114890; -.
DR   GO; GO:0005887; C:integral to plasma membrane; TAS.
DR   InterPro; IPR007110; Ig-like.
DR   Pfam; PF00047; ig; 6.
DR   PROSITE; PS50835; IG_LIKE; 6.
KW   3D-structure; Glycoprotein; GPI-anchor; Immunoglobulin domain;
KW   Lipoprotein; Membrane; Repeat; Signal.
FT   SIGNAL        1     34
FT   CHAIN        35    685       Carcinoembryonic antigen-related cell
FT                                 adhesion molecule 5.
FT   PROPEP      686    702       Removed in mature form (Potential).
FT   DOMAIN       35    144       Ig-like 1.
FT   DOMAIN      146    237       Ig-like 2.
FT   DOMAIN      238    322       Ig-like 3.
FT   DOMAIN      324    415       Ig-like 4.
FT   DOMAIN      416    498       Ig-like 5.
FT   DOMAIN      502    593       Ig-like 6.
FT   DOMAIN      594    677       Ig-like 7.
FT   CARBOHYD    104    104       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    115    115       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    152    152       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    182    182       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    197    197       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    204    204       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    208    208       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    246    246       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    256    256       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    274    274       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    288    288       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    292    292       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    309    309       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    330    330       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    351    351       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    360    360       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    375    375       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    432    432       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    466    466       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    480    480       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    508    508       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    529    529       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    553    553       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    560    560       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    580    580       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    612    612       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    650    650       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    665    665       N-linked (GlcNAc. . .) (Potential).
FT   LIPID       685    685       GPI-anchor amidated alanine (Potential).
FT   CONFLICT    320    320       Missing (in Ref. 4).
SQ   SEQUENCE   702 AA;  76795 MW;  6299AE26CDDBDB5C CRC64;

  Query Match             84.4%;  Score 38;  DB 1;  Length 702;
  Best Local Similarity   77.8%;  Pred. No. 81;
  Matches    7;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 YLSGADINL 9
              |||||::||
Db        605 YLSGANLNL 613


RESULT 3
Q8N4D0
ID   Q8N4D0      PRELIMINARY;      PRT;   702 AA.
AC   Q8N4D0;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   CEACAM5 protein.
OS   Homo sapiens (Human).
```

```
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Colon;
RX   MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA   Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
RA   Jones S.J., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length human
RT   and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Colon;
RA   Strausberg R.;
RL   Submitted (JUL-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC034671; AAH34671.1; -.
DR   HSSP; Q61353; 1L6Z.
DR   InterPro; IPR001589; Actbind_actnin.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR003598; Ig_c2.
DR   Pfam; PF00047; ig; 6.
DR   SMART; SM00408; IGc2; 3.
DR   PROSITE; PS00019; ACTININ_1; UNKNOWN_3.
DR   PROSITE; PS50835; IG_LIKE; 6.
SQ   SEQUENCE   702 AA;  76781 MW;  97CCFB7399A0B05A CRC64;

  Query Match             84.4%;  Score 38;  DB 2;  Length 702;
  Best Local Similarity   77.8%;  Pred. No. 81;
  Matches    7;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 YLSGADINL 9
              |||||::||
Db        605 YLSGANLNL 613


RESULT 4
PEX3_PICPA
ID   PEX3_PICPA     STANDARD;      PRT;   455 AA.
AC   Q92262;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   25-OCT-2004 (Rel. 45, Last annotation update)
DE   Peroxisomal biogenesis factor 3 (Peroxin-3) (Peroxisomal membrane
DE   protein PAS2).
GN   Name=PEX3; Synonyms=PAS2;
OS   Pichia pastoris (Yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Pichia.
OX   NCBI_TaxID=4922;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97115764; PubMed=8955066;
RA   Subramani S.;
RT   "Protein translocation into peroxisomes.";
RL   J. Biol. Chem. 271:32483-32486(1996).
CC   -!- FUNCTION: Involved in peroxisome biosynthesis.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein. Peroxisomal.
CC   -!- SIMILARITY: Belongs to the peroxin 3 family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z72390; CAA96530.1; -.
DR   InterPro; IPR006966; Peroxin-3.
DR   Pfam; PF04882; Peroxin-3; 1.
KW   Peroxisome; Transmembrane.
FT   DOMAIN        1     15       Peroxisomal (Potential).
FT   TRANSMEM     16     33       Potential.
FT   DOMAIN       34    455       Cytoplasmic (Potential).
SQ   SEQUENCE   455 AA;  51973 MW;  6853C5EA5C67EC34 CRC64;

  Query Match             80.0%;  Score 36;  DB 1;  Length 455;
  Best Local Similarity   77.8%;  Pred. No. 1.4e+02;
  Matches    7;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 YLSGADINL 9
              ||: |||||
Db        353 YLNNADINL 361


RESULT 5
Q8A067
ID   Q8A067      PRELIMINARY;      PRT;   322 AA.
AC   Q8A067;
DT   01-JUN-2003 (TrEMBLrel. 24, Created)
DT   01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Chitin deacetylase.
GN   OrderedLocusNames=BT4154;
OS   Bacteroides thetaiotaomicron.
OC   Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC   Bacteroidaceae; Bacteroides.
OX   NCBI_TaxID=818;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=VPI-5482 / ATCC 29148;
RX   MEDLINE=22550858; PubMed=12663928; DOI=10.1126/science.1080029;
RA   Xu J., Bjursell M.K., Himrod J., Deng S., Carmichael L.K.,
RA   Chiang H.C., Hooper L.V., Gordon J.I.;
RT   "A genomic view of the human-Bacteroides thetaiotaomicron symbiosis.";
RL   Science 299:2074-2076(2003).
DR   EMBL; AE016944; AAO79259.1; -.
DR   GO; GO:0016810; F:hydrolase activity, acting on carbon-nitrog. . .; IEA.
DR   GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR   InterPro; IPR002509; Polysac_deacet.
DR   Pfam; PF01522; Polysacc_deac_1; 1.
KW   Complete proteome.
SQ   SEQUENCE   322 AA;  37211 MW;  8A5A40FEFFBB1145 CRC64;

  Query Match             77.8%;  Score 35;  DB 2;  Length 322;
  Best Local Similarity   77.8%;  Pred. No. 1.5e+02;
  Matches    7;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 YLSGADINL 9
              |||||| |:
Db         17 YLSGADWNV 25


RESULT 6
Q65258
ID   Q65258      PRELIMINARY;      PRT;   352 AA.
AC   Q65258;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
```

```
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   ORF 13R.
OS   African swine fever virus (ASFV).
OC   Viruses; dsDNA viruses, no RNA stage; Asfarviridae; Asfivirus.
OX   NCBI_TaxID=10497;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94014996; PubMed=8409937;
RA   Vydelingum S., Baylis S.A., Bristow C., Smith G.L., Dixon L.K.;
RT   "Duplicated genes within the variable right end of the genome of a
RT   pathogenic isolate of African swine fever virus.";
RL   J. Gen. Virol. 74:2125-2130(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Dixon L.K., Twigg S.R.F., Baylis S.A., Vydelingum S., Bristow C.,
RA   Hammond J.M., Smith G.L.;
RT   "Nucleotide sequence of a 55 kbp region from the right end of the
RT   genome of a pathogenic African swine fever virus isolate (Malawi
RT   LIL20/1).";
RL   J. Gen. Virol. 7:1655-1684(1994).
DR   EMBL; X71982; CAA50855.1; -.
DR   InterPro; IPR002595; ASFV_360.
DR   Pfam; PF01671; ASFV_360; 1.
DR   ProDom; PD003462; ASFV_360; 1.
SQ   SEQUENCE   352 AA;  40682 MW;  61561D08AE1C1599 CRC64;

  Query Match             77.8%;  Score 35;  DB 2;  Length 352;
  Best Local Similarity   87.5%;  Pred. No. 1.7e+02;
  Matches    7;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 YLSGADIN 8
              || |||||
Db        258 YLLGADIN 265


RESULT 7
Q948C5
ID   Q948C5      PRELIMINARY;      PRT;   388 AA.
AC   Q948C5;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE   Hypothetical protein OSJNBa0034A02.6 (Hypothetical protein
DE   OSJNAa0029P06.15).
GN   Name=OSJNBa0034A02.6; Synonyms=OSJNAa0029P06.15;
OS   Oryza sativa (Rice).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC   Ehrhartoideae; Oryzeae; Oryza.
OX   NCBI_TaxID=4530;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Wing R.A., Yu Y., Soderlund C., Chen M., Kim H.-R., Rambo T.,
RA   Saski C., Henry D., Oates R., Simmons J.;
RL   Submitted (FEB-2002) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Wing R.A., Yu Y., Yang T.J., Nah G., Soderlund C., Chen M., Kim H.-R.,
RA   Rambo T., Saski C., Henry D., Oates R., Simmons J.;
RL   Submitted (APR-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC079852; AAL25173.1; -.
DR   EMBL; AC112513; AAM08429.1; -.
DR   Gramene; Q948C5; -.
DR   InterPro; IPR001810; F-box.
DR   InterPro; IPR011043; Gal_oxid_central.
DR   Pfam; PF00646; F-box; 1.
DR   SMART; SM00256; FBOX; 1.
DR   PROSITE; PS50181; FBOX; 1.
KW   Hypothetical protein.
SQ   SEQUENCE   388 AA;  43597 MW;  F7845EF907323CD9 CRC64;

  Query Match             77.8%;  Score 35;  DB 2;  Length 388;
  Best Local Similarity   66.7%;  Pred. No. 1.9e+02;
  Matches    6;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 YLSGADINL 9
              | :|||||:
Db        363 YTTGADINM 371


RESULT 8
Q7XGY0
ID   Q7XGY0      PRELIMINARY;      PRT;   388 AA.
AC   Q7XGY0;
DT   01-OCT-2003 (TrEMBLrel. 25, Created)
DT   01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Hypothetical protein.
GN   ORFNames=OSJNAa0029P06.15;
OS   Oryza sativa (japonica cultivar-group).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC   Ehrhartoideae; Oryzeae; Oryza.
OX   NCBI_TaxID=39947;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   The Rice Chromosome 10 Sequencing Consortium;
RT   "In-depth view of structure, activity, and evolution of rice
RT   chromosome 10.";
RL   Science 300:1566-1569(2003).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Buell C.R., Wing R.A., McCombie W.R., Messing J., Yuan Q.;
RL   Submitted (MAY-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AE017055; AAP52077.1; -.
DR   Gramene; Q7XGY0; -.
DR   InterPro; IPR001810; F-box.
DR   InterPro; IPR011043; Gal_oxid_central.
DR   Pfam; PF00646; F-box; 1.
DR   PROSITE; PS50181; FBOX; 1.
KW   Hypothetical protein.
SQ   SEQUENCE   388 AA;  43597 MW;  F7845EF907323CD9 CRC64;

  Query Match             77.8%;  Score 35;  DB 2;  Length 388;
  Best Local Similarity   66.7%;  Pred. No. 1.9e+02;
  Matches    6;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 YLSGADINL 9
              | :|||||:
Db        363 YTTGADINM 371


RESULT 9
Q6PW77
ID   Q6PW77      PRELIMINARY;      PRT;   499 AA.
AC   Q6PW77;
DT   05-JUL-2004 (TrEMBLrel. 27, Created)
DT   05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT   05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE   Glucooligosaccharide oxidase.
OS   Acremonium strictum (Black bundle disease fungus).
OC   Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC   Hypocreomycetidae; Hypocreales; Hypocreaceae; mitosporic Hypocreaceae;
OC   Acremonium.
OX   NCBI_TaxID=5046;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Lee M.-H., Lai W.-L., Lin S.-F., Liaw S.-H., Tsai Y.-C.;
RL   Submitted (MAR-2004) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY573966; AAS79317.1; -.
DR   GO; GO:0006118; P:electron transport; IEA.
DR   InterPro; IPR006094; Oxid_FAD_bind_N.
DR   InterPro; IPR006093; Oxred_FAD_BS.
DR   Pfam; PF01565; FAD_binding_4; 1.
DR   PROSITE; PS00862; OX2_COVAL_FAD; UNKNOWN_1.
SQ   SEQUENCE   499 AA;  55237 MW;  BCB56CF1F4E922CE CRC64;
```

```
  Query Match             77.8%;  Score 35;  DB 2;  Length 499;
  Best Local Similarity   66.7%;  Pred. No. 2.4e+02;
  Matches    6;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 YLSGADINL 9
              || |||:|:
Db        323 YLYGADLNI 331


RESULT 10
Q7NMP7
ID   Q7NMP7      PRELIMINARY;      PRT;   856 AA.
AC   Q7NMP7;
DT   01-MAR-2004 (TrEMBLrel. 26, Created)
DT   01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Two-component hybrid sensor and regulator.
GN   OrderedLocusNames=glr0718;
OS   Gloeobacter violaceus.
OC   Bacteria; Cyanobacteria; Chroococcales; Gloeobacter.
OX   NCBI_TaxID=33072;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=PCC 7421;
RX   MEDLINE=22977040; PubMed=14621292;
RA   Nakamura Y., Kaneko T., Sato S., Mimuro M., Miyashita H., Tsuchiya T.,
RA   Sasamoto S., Watanabe A., Kawashima K., Kishida Y., Kiyokawa C.,
RA   Kohara M., Matsumoto M., Matsuno A., Nakazaki N., Shimpo S.,
RA   Takeuchi C., Yamada M., Tabata S.;
RT   "Complete genome structure of Gloeobacter violaceus PCC 7421, a
RT   cyanobacterium that lacks thylakoids.";
RL   DNA Res. 10:137-145(2003).
CC   -!- SIMILARITY: Contains 1 histidine kinase domain.
DR   EMBL; AP006570; BAC88659.1; -.
DR   HSSP; P39928; 1OXK.
DR   GO; GO:0016020; C:membrane; IEA.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0003677; F:DNA binding; IEA.
DR   GO; GO:0016301; F:kinase activity; IEA.
DR   GO; GO:0016740; F:transferase activity; IEA.
DR   GO; GO:0000156; F:two-component response regulator activity; IEA.
DR   GO; GO:0000155; F:two-component sensor molecule activity; IEA.
DR   GO; GO:0007600; P:sensory perception; IEA.
DR   GO; GO:0000160; P:two-component signal transduction system (p. . .; IEA.
DR   InterPro; IPR003594; ATPbind_ATPase.
DR   InterPro; IPR004358; Bact_sens_pr_C.
DR   InterPro; IPR011006; CheY_like.
DR   InterPro; IPR003018; GAF.
DR   InterPro; IPR005467; His_kinase.
DR   InterPro; IPR003661; His_kinA_N.
DR   InterPro; IPR009082; His_kin_homodim.
DR   InterPro; IPR001610; PAC.
DR   InterPro; IPR000014; PAS.
DR   InterPro; IPR000700; PAS-assoc_C.
DR   InterPro; IPR001789; Response_reg.
DR   Pfam; PF01590; GAF; 1.
DR   Pfam; PF02518; HATPase_c; 1.
DR   Pfam; PF00512; HisKA; 1.
DR   Pfam; PF00785; PAC; 1.
DR   Pfam; PF00072; Response_reg; 1.
DR   PRINTS; PR00344; BCTRLSENSOR.
DR   ProDom; PD000039; Response_reg; 1.
DR   TIGRFAMs; TIGR00229; sensory_box; 1.
DR   PROSITE; PS50109; HIS_KIN; 1.
DR   PROSITE; PS50113; PAC; 1.
DR   PROSITE; PS50110; RESPONSE_REGULATORY; 1.
KW   Complete proteome; Kinase; Phosphorylation; Sensory transduction;
KW   Transferase.
SQ   SEQUENCE   856 AA;  94436 MW;  1D0C91229B3CAEF1 CRC64;

  Query Match             77.8%;  Score 35;  DB 2;  Length 856;
  Best Local Similarity   66.7%;  Pred. No. 3.9e+02;
  Matches    6;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 YLSGADINL 9
              ||:|||::|
Db         98 YLNGADLHL 106


RESULT 11
Q6ZWE8
ID   Q6ZWE8      PRELIMINARY;      PRT;   131 AA.
AC   Q6ZWE8;
DT   05-JUL-2004 (TrEMBLrel. 27, Created)
DT   05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT   05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE   Hypothetical protein FLJ41203.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Cerebellum;
RA   Tashiro H., Yamazaki M., Watanabe K., Kumagai A., Itakura S.,
RA   Fukuzumi Y., Fujimori Y., Komiyama M., Sugiyama T., Irie R.,
RA   Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J., Isono Y.,
RA   Kawai-Hio Y., Saito K., Nishikawa T., Kimura K., Yamashita H.,
RA   Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K., Wagatsuma M.,
RA   Murakawa K., Kanehori K., Takahashi-Fujii A., Oshima A., Sugiyama A.,
RA   Kawakami B., Suzuki Y., Sugano S., Nagahari K., Masuho Y., Nagai K.,
RA   Isogai T.;
RL   Submitted (JUL-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AK123197; BAC85555.1; -.
DR   HSSP; P20749; 1K1B.
DR   InterPro; IPR002110; ANK.
DR   Pfam; PF00023; Ank; 3.
DR   PRINTS; PR01415; ANKYRIN.
DR   SMART; SM00248; ANK; 3.
DR   PROSITE; PS50088; ANK_REPEAT; 3.
DR   PROSITE; PS50297; ANK_REP_REGION; 1.
KW   ANK repeat.
SQ   SEQUENCE   131 AA;  14099 MW;  FF5B3110DD406884 CRC64;

  Query Match             75.6%;  Score 34;  DB 2;  Length 131;
  Best Local Similarity   87.5%;  Pred. No. 1e+02;
  Matches    7;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          2 LSGADINL 9
              | ||||||
Db         17 LQGADINL 24


RESULT 12
Q857N2
ID   Q857N2      PRELIMINARY;      PRT;   152 AA.
AC   Q857N2;
DT   01-JUN-2003 (TrEMBLrel. 24, Created)
DT   01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Gp132.
GN   Name=132;
OS   Mycobacteriophage CJW1.
OC   Viruses; dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae.
OX   NCBI_TaxID=205869;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=22592660; PubMed=12705866; DOI=10.1016/S0092-8674(03)00233-2;
RA   Pedulla M.L., Ford M.E., Houtz J.M., Karthikeyan T., Wadsworth C.,
RA   Lewis J.A., Jacobs-Sera D., Falbo J., Gross J., Pannunzio N.R.,
RA   Brucker W., Kumar V., Kandasamy J., Keenan L., Bardarov S.,
RA   Kriakov J., Lawrence J.G., Jacobs W.R. Jr., Hendrix R.W.,
RA   Hatfull G.F.;
RT   "Origins of highly mosaic mycobacteriophage genomes.";
RL   Cell 113:171-182(2003).
```

```
DR   EMBL; AY129331; AAN01744.1; -.
SQ   SEQUENCE   152 AA;  16995 MW;  5BF89A3CFF32CA25 CRC64;

  Query Match             75.6%;  Score 34;  DB 2;  Length 152;
  Best Local Similarity   100.0%;  Pred. No. 1.2e+02;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 LSGADIN 8
              |||||||
Db         16 LSGADIN 22


RESULT 13
O58971
ID   O58971      PRELIMINARY;      PRT;   176 AA.
AC   O58971;
DT   01-AUG-1998 (TrEMBLrel. 07, Created)
DT   01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   176aa long hypothetical thermonuclease.
GN   OrderedLocusNames=PH1212;
OS   Pyrococcus horikoshii.
OC   Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
OC   Pyrococcus.
OX   NCBI_TaxID=53953;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=OT3;
RX   MEDLINE=98344137; PubMed=9679194;
RA   Kawarabayasi Y., Sawada M., Horikawa H., Haikawa Y., Hino Y.,
RA   Yamamoto S., Sekine M., Baba S.-I., Kosugi H., Hosoyama A., Nagai Y.,
RA   Sakai M., Ogura K., Otsuka R., Nakazawa H., Takamiya M., Ohfuku Y.,
RA   Funahashi T., Tanaka T., Kudoh Y., Yamazaki J., Kushida N., Oguchi A.,
RA   Aoki K.-I., Yoshizawa T., Nakamura Y., Robb F.T., Horikoshi K.,
RA   Masuchi Y., Shizuya H., Kikuchi H.;
RT   "Complete sequence and gene organization of the genome of a hyper-
RT   thermophilic archaebacterium, Pyrococcus horikoshii OT3.";
RL   DNA Res. 5:55-76(1998).
DR   EMBL; AP000005; BAA30312.1; -.
DR   PIR; F71064; F71064.
DR   HSSP; P00644; 1SNP.
DR   GO; GO:0004518; F:nuclease activity; IEA.
DR   GO; GO:0003676; F:nucleic acid binding; IEA.
DR   InterPro; IPR006021; SNase.
DR   InterPro; IPR002071; Thermonucl_AS.
DR   Pfam; PF00565; SNase; 1.
DR   SMART; SM00318; SNc; 1.
DR   PROSITE; PS01123; TNASE_1; UNKNOWN_1.
DR   PROSITE; PS50830; TNASE_3; 1.
KW   Complete proteome.
SQ   SEQUENCE   176 AA;  20010 MW;  CB802A467B17E29E CRC64;

  Query Match             75.6%;  Score 34;  DB 2;  Length 176;
  Best Local Similarity   75.0%;  Pred. No. 1.4e+02;
  Matches    6;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 YLSGADIN 8
              ||:| |||
Db        128 YLNGTDIN 135


RESULT 14
Q7U4R6
ID   Q7U4R6      PRELIMINARY;      PRT;   237 AA.
AC   Q7U4R6;
DT   01-OCT-2003 (TrEMBLrel. 25, Created)
DT   01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Hypothetical protein.
GN   OrderedLocusNames=SYNW1998;
OS   Synechococcus sp. (strain WH8102).
OC   Bacteria; Cyanobacteria; Chroococcales; Synechococcus.
OX   NCBI_TaxID=84588;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=22825697; PubMed=12917641; DOI=10.1038/nature01943;
RA   Palenik B., Brahamsha B., Larimer F.W., Land M.L., Hauser L.,
RA   Chain P., Lamerdin J.E., Regala W., Allen E.E., McCarren J.,
RA   Paulsen I.T., Dufresne A., Partensky F., Webb E.A., Waterbury J.;
RT   "The genome of a motile marine Synechococcus.";
RL   Nature 424:1037-1042(2003).
DR   EMBL; BX569694; CAE08513.1; -.
DR   InterPro; IPR001646; 5peptide_repeat.
DR   Pfam; PF00805; Pentapeptide; 3.
KW   Complete proteome; Hypothetical protein.
SQ   SEQUENCE   237 AA;  24963 MW;  9D7EF9A986A434FB CRC64;

  Query Match             75.6%;  Score 34;  DB 2;  Length 237;
  Best Local Similarity   75.0%;  Pred. No. 1.8e+02;
  Matches    6;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 YLSGADIN 8
              ||||||::
Db        125 YLSGADLS 132


RESULT 15
Q96LU9
ID   Q96LU9      PRELIMINARY;      PRT;   268 AA.
AC   Q96LU9;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Hypothetical protein FLJ25053.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Cerebellum;
RA   Nishi T., Nakagawa S., Senoh A., Mizuguchi H., Inagaki H., Suzuki Y.,
RA   Hata H., Nakagawa K., Mizuno S., Morinaga M., Kawamura M.,
RA   Sugiyama T., Irie R., Otsuki T., Sato H., Nishikawa T., Sugiyama A.,
RA   Kawakami B., Nagai K., Isogai T., Sugano S.;
RL   Submitted (OCT-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AK057782; BAB71569.1; -.
DR   HSSP; P42773; 1MX4.
DR   InterPro; IPR002110; ANK.
DR   Pfam; PF00023; Ank; 7.
DR   PRINTS; PR01415; ANKYRIN.
DR   SMART; SM00248; ANK; 7.
DR   PROSITE; PS50088; ANK_REPEAT; 5.
DR   PROSITE; PS50297; ANK_REP_REGION; 1.
KW   ANK repeat.
SQ   SEQUENCE   268 AA;  29157 MW;  5CBF876C344CF3F6 CRC64;

  Query Match             75.6%;  Score 34;  DB 2;  Length 268;
  Best Local Similarity   87.5%;  Pred. No. 2.1e+02;
  Matches    7;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          2 LSGADINL 9
              | ||||||
Db         67 LQGADINL 74


Search completed: May 17, 2005, 06:23:36
Job time : 53.75 secs
```

GenCore version 5.1.6

OM protein - nucleic search, using frame_plus_p2n model

Run on: May 17, 2005, 16:29:39 ; Search time 326.5 Seconds
(without alignments)
163.178 Million cell updates/sec

Title: US-10-725-373-5
Perfect score: 48
Sequence: 1 YLSGACLNL 9

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/US10725373/runat_17052005_071020_16177/app_query.fasta_1.796
-DB=N_Geneseq_16Dec04 -QFMT=fastap -SUFFIX=rng -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10725373_@CGN_1_1_1241_@runat_17052005_071020_16177 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N_Geneseq_16Dec04:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002as:*
7: geneseqn2002bs:*
8: geneseqn2003as:*
9: geneseqn2003bs:*
10: geneseqn2003cs:*
11: geneseqn2003ds:*
12: geneseqn2004as:*
13: geneseqn2004bs:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	42	87.5	110000	13	ABD32780_3	Continuation (4 of
2	41	85.4	1884	10	ADF60975	Adf60975 B. thurin
3	41	85.4	1899	4	AAS02464	Aas02464 B. thurin
4	41	85.4	1899	10	ADF60973	Adf60973 B. thurin
5	41	85.4	1899	10	ADF60971	Adf60971 B. thurin
6	41	85.4	1902	2	AAV52612	Aav52612 Nucleotid
7	41	85.4	1902	2	AAV52611	Aav52611 Nucleotid
8	41	85.4	1912	3	AAA15566	Aaa15566 Cry2Ab de
9	41	85.4	2924	2	AAQ71027	Aaq71027 CryIIB ge
c 10	39	81.2	822	4	AAH52607	Aah52607 S. epider
11	39	81.2	1032	13	ADS45359	Ads45359 Bacterial
12	39	81.2	1384	3	AAC74312	Aac74312 Human sec
13	39	81.2	2950	4	AAH53985	Aah53985 S. epider
14	39	81.2	3760	4	AAH54665	Aah54665 S. epider
c 15	39	81.2	41202	12	ADQ97382	Adq97382 Mouse can
16	39	81.2	80332	11	ACN44842	Acn44842 Human gen
17	38	79.2	374	8	ABZ19661	Abz19661 Group III
18	38	79.2	610	13	ADQ57642	Adq57642 Novel can
19	38	79.2	848	12	ADJ67429	Adj67429 Human ova
c 20	38	79.2	862	12	ADJ67428	Adj67428 Human ova
21	38	79.2	1509	11	ABD00239	Abd00239 Klebsiell
22	38	79.2	2540	4	ABL26320	Abl26320 Drosophil
c 23	38	79.2	5002	12	ADQ64368	Adq64368 Novel hum
c 24	38	79.2	8801	4	ABL28642	Abl28642 Drosophil
25	38	79.2	10793	4	ABL26318	Abl26318 Drosophil
c 26	38	79.2	42999	6	ABK90832	Abk90832 Genomic D
c 27	38	79.2	51558	13	ACN37207	Acn37207 Human per
28	38	79.2	110000	6	ABX08336_04	Continuation (5 of
29	38	79.2	110000	12	ADJ25985_04	Continuation (5 of
30	38	79.2	110000	12	ADN97989_04	Continuation (5 of
31	38	79.2	110000	12	ADO50281_04	Continuation (5 of
c 32	38	79.2	114411	12	ADQ21090	Adq21090 Human sof
33	37	77.1	250	4	AAK69831	Aak69831 Human imm
34	37	77.1	300	4	AAK69830	Aak69830 Human imm
35	37	77.1	300	4	AAK57233	Aak57233 Human imm
c 36	37	77.1	354	6	ABN18754	Abn18754 Human ORF
37	37	77:1	375	8	ABZ18602	Abz18602 Group III
38	37	77.1	401	4	AAK96284	Aak96284 Human neu
39	37	77.1	401	4	AAK97777	Aak97777 Human neu
40	37	77.1	401	6	ABT01054	Abt01054 Human neu
41	37	77.1	401	6	ABT02547	Abt02547 Human neu
42	37	77.1	433	6	ABV99276	Abv99276 Marine sn
c 43	37	77.1	499	4	AAK57892	Aak57892 Human imm
c 44	37	77.1	503	9	ACH34230	Ach34230 Human end
c 45	37	77.1	672	12	ADQ20409	Adq20409 Human sof

ALIGNMENTS

RESULT 1
ABD32780_3
Continuation (4 of 5) of ABD32780 from base 300001 (Human cancer-associated genomic DNA F
WP Sequence split into 5 fragments LOCUS ABD32780 Accession Abd32780

WP	Fragment Name	Begin	End
WP	ABD32780_0	1	110000
WP	ABD32780_1	100001	210000
WP	ABD32780_2	200001	310000
WP	ABD32780_3	300001	410000
WP	ABD32780_4	400001	430442

Alignment Scores:

Pred. No.:	2.77e+04	Length:	110000
Score:	42.00	Matches:	8
Percent Similarity:	88.89%	Conservative:	0
Best Local Similarity:	88.89%	Mismatches:	1
Query Match:	87.50%	Indels:	0
DB:	13	Gaps:	0

US-10-725-373-5 (1-9) x ABD32780_3 (1-110000)

```
Qy          1 TyrLeuSerGlyAlaCysLeuAsnLeu 9
              ||||||||||||   ||||||||||||
Db      47157 TATCTGTCTGGGCATTGTCTTAACCTC 47183
```

RESULT 2
ADF60975
ID ADF60975 standard; DNA; 1884 BP.
XX

```
AC   ADF60975;
XX
DT   12-FEB-2004  (first entry)
XX
DE   B. thuringiensis Cry2Ag DNA.
XX
KW   Cry2Ag; gene; ds; insecticidal protein; insecticide; insect control;
KW   insect damage; plant resistance.
XX
OS   Bacillus thuringiensis.
XX
FH   Key             Location/Qualifiers
FT   CDS             1. .1884
FT                   /*tag=  a
FT                   /product= "B. thuringiensis Cry2Ag"
XX
PN   US2003167517-A1.
XX
PD   04-SEP-2003.
XX
PF   09-JAN-2002; 2002US-00040906.
XX
PR   09-JAN-2001; 2001US-00331355.
XX
PA   (ARNA/) ARNAUT G.
PA   (BOET/) BOETS A.
PA   (VANN/) VANNESTE S.
PA   (VRIE/) VAN RIE J.
PA   (VHOU/) VAN HOUDT S.
XX
PI   Arnaut G,  Boets A,  Vanneste S,  Van Rie J,  Van Houdt S;
XX
DR   WPI; 2003-898134/82.
DR   P-PSDB; ADF60976.
XX
PT   New Cry2Ae, Cry2Af and Cry2Ag insecticidal proteins, useful for
PT   protecting plants from insect damage, for controlling insects, or for
PT   rendering a plant resistant to an insect.
XX
PS   Claim 7; SEQ ID NO 5; 32pp; English.
XX
CC   The invention relates to insecticidal proteins, designated Cry2Ae, Cry2Af
CC   and Cry2Ag. The invention also relates to nucleic acid sequences encoding
CC   the insecticidal Cry2Ae, Cry2Af and Cry2Ag proteins, a chimeric gene
CC   comprising one of the nucleic acid sequences under the control of a plant
CC   -expressible promoter, plant cells, plants or seeds transformed to
CC   comprise the chimeric gene, a microorganism transformed to comprise the
CC   nucleic acid sequence, a process for rendering a plant resistant to an
CC   insect comprising transforming plant cells with the chimeric gene and
CC   regenerating transformed plants from the cells that are resistant to
CC   insects, and a method for controlling insects comprising expressing the
CC   Cry2Ae, Cry2Af and Cry2Ag proteins in transformed plant cells. The
CC   Cry2Ae, Cry2Af and Cry2Ag proteins and nucleic acids encoding the
CC   proteins are useful for protecting plants from insect damage, for
CC   controlling insects or for rendering a plant resistant to an insect. This
CC   sequence represents DNA encoding the Cry2Ag protein of the invention.
XX
SQ   Sequence 1884 BP; 638 A; 292 C; 348 G; 606 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:              408          Length:          1884
Score:                  41.00        Matches:         7
Percent Similarity:     88.89%       Conservative:    1
Best Local Similarity:  77.78%       Mismatches:      1
Query Match:            85.42%       Indels:          0
DB:                     10           Gaps:            0

US-10-725-373-5 (1-9) x ADF60975 (1-1884)

Qy          1 TyrLeuSerGlyAlaCysLeuAsnLeu 9
              ||||||||||||   |||||||||:::
Db        767 TATCTATCTGGTCGTTGTTTAAATATC 793
```

```
RESULT 3
AAS02464
ID   AAS02464 standard; DNA; 1899 BP.
XX
AC   AAS02464;
XX
DT   29-AUG-2001  (first entry)
XX
DE   B. thuringiensis DNA encoding a toxic crystal protein, CryET31.
XX
KW   Delta endotoxin; Lepidopteran-active; crystal protein; insecticide;
KW   transgenic plant; corn; wheat; soybean; oat; cotton; rice; rye; sorghum;
KW   sugarcane; tomato; tobacco; kapok; flax; potato; barley; turf grass;
KW   pasture grass; berry; fruit; legume; vegetable; ornamental plant; shrub;
KW   cactus; tree cell; gypsy moth; looper; tobacco budworm; spruce budworm;
KW   cotton leaf perforator; CryET31; ds.
XX
OS   Bacillus thuringiensis.
XX
FH   Key             Location/Qualifiers
FT   CDS             1. .1899
FT                   /*tag=  a
FT                   /product= "CryET31"
XX
PN   WO200119859-A2.
XX
PD   22-MAR-2001.
XX
PF   13-SEP-2000; 2000WO-US025361.
XX
PR   15-SEP-1999;   99US-0153995P.
XX
PA   (MONS ) MONSANTO CO.
XX
PI   Baum JA,  Chu C,  Donovan WP,  Gilmer AJ,  Rupar MJ;
XX
DR   WPI; 2001-281518/29.
DR   P-PSDB; AAU02021.
XX
PT   Lepidopteran-active Bacillus thuringiensis delta-endotoxin polypeptides
PT   and the polynucleotides that encode them, useful for increasing the
PT   insect resistance of plant.
XX
PS   Claim 17; Page 99-102; 173pp; English.
XX
CC   The sequence encodes a B. thuringiensis Lepidopteran-active delta-
CC   endotoxin, crystal protein CryET31. The Lepidopteran-active B.
CC   thuringiensis delta-endotoxin polypeptides may be used as compositions
CC   that are applied to plant crops to protect them from insect damage. The
CC   polynucleotides may be used in the production of transgenic plants that
CC   express the insecticidal polypeptides and consequently have improved
CC   insect resistance compared to non-transformed plants. Monocotyledonous or
CC   dicotyledonous plants may be protected in this way, for example corn,
CC   wheat, soybean, oat, cotton, rice, rye, sorghum, sugarcane, tomato,
CC   tobacco, kapok, flax, potato, barley, turf grass, pasture grass, berry,
CC   fruit, legume, vegetable, ornamental plant, shrub, cactus and/or tree
CC   cell. A wide range of insects (e.g. gypsy moth, looper, tobacco budworm,
CC   cotton leaf perforator and spruce budworm) may be affected by application
CC   of the insecticidal polypeptides (full details given in specification)
XX
SQ   Sequence 1899 BP; 619 A; 303 C; 356 G; 621 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:              412          Length:          1899
Score:                  41.00        Matches:         7
Percent Similarity:     88.89%       Conservative:    1
Best Local Similarity:  77.78%       Mismatches:      1
Query Match:            85.42%       Indels:          0
DB:                     4            Gaps:            0

US-10-725-373-5 (1-9) x AAS02464 (1-1899)

Qy          1 TyrLeuSerGlyAlaCysLeuAsnLeu 9
```

```
                  |||||||||||   |||||||||:::
Db        767 TATCTATCTGGTCGTTGTTTAAATATC 793

RESULT 4
ADF60973
ID   ADF60973 standard; DNA; 1899 BP.
XX
AC   ADF60973;
XX
DT   12-FEB-2004  (first entry)
XX
DE   B. thuringiensis Cry2Af DNA.
XX
KW   Cry2Af; gene; ds; insecticidal protein; insecticide; insect control;
KW   insect damage; plant resistance.
XX
OS   Bacillus thuringiensis.
XX
FH   Key             Location/Qualifiers
FT   CDS             1. .1899
FT                   /*tag=  a
FT                   /product= "B. thuringiensis Cry2Af"
XX
PN   US2003167517-A1.
XX
PD   04-SEP-2003.
XX
PF   09-JAN-2002; 2002US-00040906.
XX
PR   09-JAN-2001; 2001US-00331355.
XX
PA   (ARNA/) ARNAUT G.
PA   (BOET/) BOETS A.
PA   (VANN/) VANNESTE S.
PA   (VRIE/) VAN RIE J.
PA   (VHOU/) VAN HOUDT S.
XX
PI   Arnaut G,  Boets A,  Vanneste S,  Van Rie J,  Van Houdt S;
XX
DR   WPI; 2003-898134/82.
DR   P-PSDB; ADF60974.
XX
PT   New Cry2Ae, Cry2Af and Cry2Ag insecticidal proteins, useful for
PT   protecting plants from insect damage, for controlling insects, or for
PT   rendering a plant resistant to an insect.
XX
PS   Claim 4; SEQ ID NO 3; 32pp; English.
XX
CC   The invention relates to insecticidal proteins, designated Cry2Ae, Cry2Af
CC   and Cry2Ag. The invention also relates to nucleic acid sequences encoding
CC   the insecticidal Cry2Ae, Cry2Af and Cry2Ag proteins, a chimeric gene
CC   comprising one of the nucleic acid sequences under the control of a plant
CC   -expressible promoter, plant cells, plants or seeds transformed to
CC   comprise the chimeric gene, a microorganism transformed to comprise the
CC   nucleic acid sequence, a process for rendering a plant resistant to an
CC   insect comprising transforming plant cells with the chimeric gene and
CC   regenerating transformed plants from the cells that are resistant to
CC   insects, and a method for controlling insects comprising expressing the
CC   Cry2Ae, Cry2Af and Cry2Ag proteins in transformed plant cells. The
CC   Cry2Ae, Cry2Af and Cry2Ag proteins and nucleic acids encoding the
CC   proteins are useful for protecting plants from insect damage, for
CC   controlling insects or for rendering a plant resistant to an insect. This
CC   sequence represents DNA encoding the Cry2Af protein of the invention.
XX
SQ   Sequence 1899 BP; 624 A; 300 C; 350 G; 625 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:                412             Length:          1899
Score:                    41.00           Matches:         7
Percent Similarity:       88.89%          Conservative:    1
Best Local Similarity:    77.78%          Mismatches:      1
Query Match:              85.42%          Indels:          0
DB:                       10              Gaps:            0

US-10-725-373-5 (1-9) x ADF60973 (1-1899)

Qy          1 TyrLeuSerGlyAlaCysLeuAsnLeu 9
                |||||||||||   |||||||||:::
Db        767 TATCTATCTGGTCGTTGTTTAAATATC 793

RESULT 5
ADF60971
ID   ADF60971 standard; DNA; 1899 BP.
XX
AC   ADF60971;
XX
DT   12-FEB-2004  (first entry)
XX
DE   B. thuringiensis Cry2Ae DNA.
XX
KW   Cry2Ae; gene; ds; insecticidal protein; insecticide; insect control;
KW   insect damage; plant resistance.
XX
OS   Bacillus thuringiensis.
XX
FH   Key             Location/Qualifiers
FT   CDS             1. .1899
FT                   /*tag=  a
FT                   /product= "B. thuringiensis Cry2Ae"
XX
PN   US2003167517-A1.
XX
PD   04-SEP-2003.
XX
PF   09-JAN-2002; 2002US-00040906.
XX
PR   09-JAN-2001; 2001US-00331355.
XX
PA   (ARNA/) ARNAUT G.
PA   (BOET/) BOETS A.
PA   (VANN/) VANNESTE S.
PA   (VRIE/) VAN RIE J.
PA   (VHOU/) VAN HOUDT S.
XX
PI   Arnaut G,  Boets A,  Vanneste S,  Van Rie J,  Van Houdt S;
XX
DR   WPI; 2003-898134/82.
DR   P-PSDB; ADF60972.
XX
PT   New Cry2Ae, Cry2Af and Cry2Ag insecticidal proteins, useful for
PT   protecting plants from insect damage, for controlling insects, or for
PT   rendering a plant resistant to an insect.
XX
PS   Claim 1; SEQ ID NO 1; 32pp; English.
XX
CC   The invention relates to insecticidal proteins, designated Cry2Ae, Cry2Af
CC   and Cry2Ag. The invention also relates to nucleic acid sequences encoding
CC   the insecticidal Cry2Ae, Cry2Af and Cry2Ag proteins, a chimeric gene
CC   comprising one of the nucleic acid sequences under the control of a plant
CC   -expressible promoter, plant cells, plants or seeds transformed to
CC   comprise the chimeric gene, a microorganism transformed to comprise the
CC   nucleic acid sequence, a process for rendering a plant resistant to an
CC   insect comprising transforming plant cells with the chimeric gene and
CC   regenerating transformed plants from the cells that are resistant to
CC   insects, and a method for controlling insects comprising expressing the
CC   Cry2Ae, Cry2Af and Cry2Ag proteins in transformed plant cells. The
CC   Cry2Ae, Cry2Af and Cry2Ag proteins and nucleic acids encoding the
CC   proteins are useful for protecting plants from insect damage, for
CC   controlling insects or for rendering a plant resistant to an insect. This
CC   sequence represents DNA encoding the Cry2Ae protein of the invention.
XX
SQ   Sequence 1899 BP; 619 A; 303 C; 356 G; 621 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:                412             Length:          1899
Score:                    41.00           Matches:         7
```

```
Percent Similarity:     88.89%          Conservative:  1
Best Local Similarity:  77.78%          Mismatches:    1
Query Match:            85.42%          Indels:        0
DB:                     10              Gaps:          0

US-10-725-373-5 (1-9) x ADF60971 (1-1899)

Qy          1 TyrLeuSerGlyAlaCysLeuAsnLeu 9
              ||||||||||||   |||||||||:::
Db        767 TATCTATCTGGTCGTTGTTTAAATATC 793

RESULT 6
AAV52612
ID   AAV52612 standard; DNA; 1902 BP.
XX
AC   AAV52612;
XX
DT   02-DEC-1998  (first entry)
XX
DE   Nucleotide sequence of lepidoteran-active 86I2 toxin.
XX
KW   86I2 toxin; PCR; primer; amplification; Bacillus thuringiensis; probe;
KW   lepidoptera; pest; pesticide; Ostrinia nubilalis; Heliothis virescens;
KW   Helicoverpa zea; hybridisation; ss.
XX
OS   Bacillus thuringiensis.
XX
FH   Key             Location/Qualifiers
FT   CDS             1. .1902
FT                   /*tag=  a
FT                   /product= "86I2 toxin"
XX
PN   WO9840490-A1.
XX
PD   17-SEP-1998.
XX
PF   13-MAR-1998;   98WO-US005081.
XX
PR   13-MAR-1997;   97US-0040512P.
XX
PA   (MYCO ) MYCOGEN CORP.
XX
PI   Schnepf HE,  Narva KE,  Muller-Cohn J;
XX
DR   WPI; 1998-506734/43.
DR   P-PSDB; AAW75775.
XX
PT   New insecticidal Bacillus thuringiensis toxins - useful for controlling
PT   lepidopteran pests, especially Ostrinia nubilalis, Heliothis virescens
PT   and Helicoverpa zea.
XX
PS   Claim 12; Page 35-36; 50pp; English.
XX
CC   This is the nucleotide sequence of a novel Bacillus thuringiensis toxin
CC   used in the method of the invention, to control lepidopteran pests. The
CC   new toxins are useful as pesticides, especially for the control of
CC   Ostrinia nubilalis, Heliothis virescens, and Helicoverpa zea. The
CC   polynucleotide coding sequences are useful for recombinant expression of
CC   the toxins and the primers, together with probes derived from the new
CC   sequences, are useful for the identification and characterisation of
CC   novel genes that encode pesticidal toxins
XX
SQ   Sequence 1902 BP; 637 A; 302 C; 338 G; 625 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:              413             Length:        1902
Score:                  41.00           Matches:       7
Percent Similarity:     88.89%          Conservative:  1
Best Local Similarity:  77.78%          Mismatches:    1
Query Match:            85.42%          Indels:        0
DB:                     2               Gaps:          0

US-10-725-373-5 (1-9) x AAV52612 (1-1902)

Qy          1 TyrLeuSerGlyAlaCysLeuAsnLeu 9
              ||||||||||||   |||||||||:::
Db        767 TATCTATCTGGTCGTTGTTTAAATATC 793

RESULT 7
AAV52611
ID   AAV52611 standard; DNA; 1902 BP.
XX
AC   AAV52611;
XX
DT   02-DEC-1998  (first entry)
XX
DE   Nucleotide sequence of lepidoteran-active HD525 toxin.
XX
KW   HD525 toxin; PCR; primer; amplification; Bacillus thuringiensis; probe;
KW   lepidoptera; pest; pesticide; Ostrinia nubilalis; Heliothis virescens;
KW   Helicoverpa zea; hybridisation; ss.
XX
OS   Bacillus thuringiensis.
XX
FH   Key             Location/Qualifiers
FT   CDS             1. .1902
FT                   /*tag=  a
FT                   /product= "HD525 toxin"
XX
PN   WO9840490-A1.
XX
PD   17-SEP-1998.
XX
PF   13-MAR-1998;   98WO-US005081.
XX
PR   13-MAR-1997;   97US-0040512P.
XX
PA   (MYCO ) MYCOGEN CORP.
XX
PI   Schnepf HE,  Narva KE,  Muller-Cohn J;
XX
DR   WPI; 1998-506734/43.
DR   P-PSDB; AAW75774.
XX
PT   New insecticidal Bacillus thuringiensis toxins - useful for controlling
PT   lepidopteran pests, especially Ostrinia nubilalis, Heliothis virescens
PT   and Helicoverpa zea.
XX
PS   Claim 17; Page 31-32; 50pp; English.
XX
CC   This is the nucleotide sequence of a novel Bacillus thuringiensis toxin
CC   used in the method of the invention, to control lepidopteran pests. The
CC   new toxins are useful as pesticides, especially for the control of
CC   Ostrinia nubilalis, Heliothis virescens, and Helicoverpa zea. The
CC   polynucleotide coding sequences are useful for recombinant expression of
CC   the toxins and the primers, together with probes derived from the new
CC   sequences, are useful for the identification and characterisation of
CC   novel genes that encode pesticidal toxins
XX
SQ   Sequence 1902 BP; 633 A; 304 C; 337 G; 628 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:              413             Length:        1902
Score:                  41.00           Matches:       7
Percent Similarity:     88.89%          Conservative:  1
Best Local Similarity:  77.78%          Mismatches:    1
Query Match:            85.42%          Indels:        0
DB:                     2               Gaps:          0

US-10-725-373-5 (1-9) x AAV52611 (1-1902)

Qy          1 TyrLeuSerGlyAlaCysLeuAsnLeu 9
              ||||||||||||   |||||||||:::
Db        767 TATCTATCTGGTCGTTGTTTAAATATC 793

RESULT 8
```

```
AAA15566
ID   AAA15566 standard; DNA; 1912 BP.
XX
AC   AAA15566;
XX
DT   28-JUL-2000  (first entry)
XX
DE   Cry2Ab delta-endotoxin gene.
XX
KW   Transgenic plant; insect resistance; cry2Ab delta-endotoxin; Coleopteran;
KW   Lepidopteran; Dipteran; plastid transit peptide; PTP; insecticidal;
KW   plasid targeting peptide; ds.
XX
OS   Bacillus thuringiensis.
XX
FH   Key             Location/Qualifiers
FT   CDS             1. .1902
FT                   /*tag=  a
FT                   /product= "cry2Ab delta-endotoxin"
XX
PN   WO200026371-A1.
XX
PD   11-MAY-2000.
XX
PF   04-NOV-1999;   99WO-US026086.
XX
PR   04-NOV-1998;   98US-00186002.
XX
PA   (MONS ) MONSANTO CO.
XX
PI   Corbin DR,  Romano CP;
XX
DR   WPI; 2000-376130/32.
DR   P-PSDB; AAY94260.
XX
PT   New method of expressing insecticidal proteins in plants transformed with
PT   a Bacillus thuringiensis delta-endotoxin encoding gene resulting in
PT   effective control of susceptible target pests.
XX
PS   Claim 12; Page 99; 104pp; English.
XX
CC   The present sequence is the cry2Ab delta-endotoxin gene. Delta-endotoxins
CC   are produced by Bacillus thuringiensis during sporulation. These proteins
CC   are toxic to certain species of insect e.g. Lepidopteran and Coleopteran
CC   larvae. An insect-resistant transgenic plant has been constructed which
CC   contains the present sequence. The cry2Ab gene would be transferred into
CC   plants via expression vectors, which subsequently allow high expression
CC   of the cry2Ab gene. The present sequence lacks Dipteran inhibitory
CC   activity. Protection may be attained against insects such as Ostrina
CC   spp., Diatraea spp., Helicoverpa spp., and Spodoptera spp., in Zea mays;
CC   Heliothis virescens, Helicovera spp., Pectinophora spp., in Gossypium
CC   hirsutum; Anticarsi spp., Pseudoplusia spp., Epinotia spp., in Glycine
CC   max; and Scirpophaga incertulas in Oryza sativa. Expression of the
CC   present sequence by a plant cell produces a fusion protein comprising an
CC   amino-terminal plastid transit peptide (PTP) covalently linked to the
CC   delta-endotoxin. The fusion protein functions to localise the delta-
CC   endotoxin to a subcellular organelle or compartment
XX
SQ   Sequence 1912 BP; 627 A; 304 C; 351 G; 630 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:              415          Length:        1912
Score:                  41.00        Matches:       7
Percent Similarity:     88.89%       Conservative:  1
Best Local Similarity:  77.78%       Mismatches:    1
Query Match:            85.42%       Indels:        0
DB:                     3            Gaps:          0

US-10-725-373-5 (1-9) x AAA15566 (1-1912)

Qy          1 TyrLeuSerGlyAlaCysLeuAsnLeu 9
              ||||||||||||   |||||||||:::
Db        767 TATCTATCTGGTCGTTGTTTAAATATC 793
```

```
RESULT 9
AAQ71027
ID   AAQ71027 standard; DNA; 2924 BP.
XX
AC   AAQ71027;
XX
DT   25-MAR-2003  (revised)
DT   27-MAR-1995  (first entry)
XX
DE   CryIIB gene which encodes insecticidal crystal protein.
XX
KW   CryIIA; CryIIIA; CryIIB; CryC; P-2; CryBI; insecticidal protein crystal;
KW   lepidoptera; environmental insecticide; Bacillus thuringiensis; toxic;
KW   probe; hybridisation; ss.
XX
OS   Bacillus thuringiensis.
XX
FH   Key             Location/Qualifiers
FT   RBS             860. .865
FT                   /*tag=  a
FT   CDS             874. .2775
FT                   /*tag=  b
FT                   /product= "CryIIB protein"
XX
PN   US5338544-A.
XX
PD   16-AUG-1994.
XX
PF   26-FEB-1993;   93US-00023736.
XX
PR   16-APR-1987;   87US-00039542.
PR   11-JUL-1989;   89US-00379015.
PR   28-AUG-1991;   91US-00751452.
XX
PA   (ECOG-) ECOGEN INC.
XX
PI   Donovan WP;
XX
DR   WPI; 1994-263236/32.
DR   P-PSDB; AAR56697.
XX
PT   New Cry IIB protein - obtd. from the cry II B gene in Bacillus
PT   thuringiensis var. Kurstaki, active against lepidopteran insects.
XX
PS   Claim 2; Fig 6A-6D; 39pp; English.
XX
CC   CryIIB encodes an insecticidal crystal protein isolated from Bacillus
CC   thuringiensis var. kurstaki (B.t.k.). The CryIIA gene was used as a probe
CC   to identify clones contg. CryIIB. The CryIIB gene does not express well
CC   with its native promoter, and so a recombinant hybrid fusion gene in
CC   which the promoter from the CryIIIA gene was fused to the protein coding
CC   region of the CryIIB gene. B.t.k produces crystal proteins during
CC   sporulation which are specifically toxic to certain orders and species of
CC   insects, esp. Lepidopterans. CryIIB can be used in compositions used as
CC   environmentally acceptable insecticides. (See also AAQ71025-6) (Updated
CC   on 25-MAR-2003 to correct PF field.)
XX
SQ   Sequence 2924 BP; 991 A; 416 C; 524 G; 993 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:              676          Length:        2924
Score:                  41.00        Matches:       7
Percent Similarity:     88.89%       Conservative:  1
Best Local Similarity:  77.78%       Mismatches:    1
Query Match:            85.42%       Indels:        0
DB:                     2            Gaps:          0

US-10-725-373-5 (1-9) x AAQ71027 (1-2924)

Qy          1 TyrLeuSerGlyAlaCysLeuAsnLeu 9
              ||||||||||||   |||||||||:::
Db       1640 TATCTATCTGGTCGTTGTTTAAATATC 1666
```

```
RESULT 10
AAH52607/c
ID   AAH52607 standard; DNA; 822 BP.
XX
AC   AAH52607;
XX
DT   03-SEP-2001  (first entry)
XX
DE   S. epidermidis open reading frame nucleotide sequence SEQ ID NO:607.
XX
KW   Staphylococcus epidermidis SR1 strain; infection; diagnosis; vaccination;
KW   endocarditis; ds.
XX
OS   Staphylococcus epidermidis.
XX
PN   WO200134809-A2.
XX
PD   17-MAY-2001.
XX
PF   09-NOV-2000; 2000WO-US030782.
XX
PR   09-NOV-1999;   99US-0164258P.
XX
PA   (GLAX ) GLAXO GROUP LTD.
XX
PI   Kimmerly WJ;
XX
DR   WPI; 2001-316495/33.
DR   P-PSDB; AAG81757.
XX
PT   Nucleic acids encoding polypeptides from Staphylococcus epidermidis,
PT   useful for vaccinating against infections, e.g. endocarditis.
XX
PS   Claim 8; Page 196; 2188pp; English.
XX
CC   AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides
CC   (II), given in AAG81454 to AAG83120, from Staphylococcus epidermidis. (I)
CC   and (II) can have antibacterial activity and therefore can be used in
CC   vaccination. The nucleic acids (I) may be used to produce the S.
CC   epidermidis polypeptides (II) via the production of vectors containing
CC   them which are used to produce hosts cells which express the
CC   polypeptides. The polypeptides (II) (and/or nucleic acids) may then be
CC   used to vaccinate subjects and to raise antibodies against the bacteria.
CC   The polypeptides may also be used to assay for other inhibitors of their
CC   activity and therefore identify compounds that may be used for the
CC   treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to
CC   AAH55090 represent specifically claimed S. epidermidis genomic DNA
CC   polynucleotide sequences from the present invention. AAH55091 to AAH55098
CC   represent oligonucleotide sequences and primers which are used in the
CC   exemplification of the present invention. N.B. The present invention
CC   specifically claims all the polynucleotide sequences given in the
CC   sequence listing of the present specification, however the sequence
CC   listing only goes up to SEQ ID NO:4454 so even though sequences are given
CC   in the disclosure for SEQ ID NO:4465 to 4472, no sequences are present
CC   for SEQ ID NO:4455 to 4464
XX
SQ   Sequence 822 BP; 342 A; 98 C; 161 G; 221 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:              394          Length:        822
Score:                  39.00        Matches:       6
Percent Similarity:     88.89%       Conservative:  2
Best Local Similarity:  66.67%       Mismatches:    1
Query Match:            81.25%       Indels:        0
DB:                     4            Gaps:          0

US-10-725-373-5 (1-9) x AAH52607 (1-822)

Qy          1 TyrLeuSerGlyAlaCysLeuAsnLeu 9
              |||   ||||||:::|||:::||||||
Db        105 TATTTCAGTGGCTCGTGCATTAATTTG 79
```

```
RESULT 11
ADS45359
ID   ADS45359 standard; cDNA; 1032 BP.
XX
AC   ADS45359;
XX
DT   02-DEC-2004  (first entry)
XX
DE   Bacterial polynucleotide #102.
XX
KW   Recombinant DNA construct; transformed plant; improved plant property;
KW   cold tolerance; heat tolerance; drought tolerance; herbicide; osmosis;
KW   pathogen tolerance; pest tolerance; plant disease resistance;
KW   cell cycle pathway modification; plant growth regulator;
KW   homologous recombination; seed oil yield; protein yield; carbohydrate;
KW   nitrogen; phosphorus; photosynthesis; lignin; galactomannan;
KW   bacterial polynucleotide; gene; ss.
XX
OS   Bacteria.
XX
PN   US2003233675-A1.
XX
PD   18-DEC-2003.
XX
PF   20-FEB-2003; 2003US-00369493.
XX
PR   21-FEB-2002; 2002US-0360039P.
XX
PA   (CAOY/) CAO Y.
PA   (HINK/) HINKLE G J.
PA   (SLAT/) SLATER S C.
PA   (CHEN/) CHEN X.
PA   (GOLD/) GOLDMAN B S.
XX
PI   Cao Y,  Hinkle GJ,  Slater SC,  Chen X,  Goldman BS;
XX
DR   WPI; 2004-061375/06.
XX
PT   New recombinant DNA construct comprising a promoter positioned to provide
PT   for expression of a polynucleotide encoding a polypeptide from a
PT   microbial source, useful for producing plants with improved properties.
XX
PS   Claim 1; SEQ ID NO 23789; 122pp; English.
XX
CC   The invention relates to a recombinant DNA construct comprising a
CC   promoter functional in a plant cell, where the promoter is positioned to
CC   provide for expression of a polynucleotide encoding a polypeptide from a
CC   microbial source. The invention also relates to a transformed plant
CC   comprising the recombinant DNA construct and a method of producing a
CC   transformed plant having an improved property. The plant is a crop plant
CC   such as maize or soybean. The method of producing a transformed plant
CC   having an improved property comprises transforming a plant with the
CC   recombinant DNA construct and growing the transformed plant, where the
CC   polynucleotide or polypeptide is useful for improving plant properties.
CC   The recombinant DNA construct is useful for producing plants with
CC   improved plant properties, e.g. improved cold, heat or drought tolerance,
CC   tolerance to herbicides, extreme osmotic conditions, pathogens or pests,
CC   increased resistance to plant disease, better growth rate by modification
CC   of the cell cycle pathway with plant growth regulators, increased rate of
CC   homologous recombination, modified seed oil or protein yield and/or
CC   content, improved yield by modification of carbohydrate, nitrogen or
CC   phosphorus use and/or uptake, by modification of photosynthesis or by
CC   providing improved plant growth and development under at least one stress
CC   condition, improved lignin production or improved galactomannan
CC   production. This sequence represents a bacterial polynucleotide used in
CC   the scope of the invention. Note: The sequence data for this patent did
CC   not form part of the printed specification but was obtained in electronic
CC   format from USPTO at seqdata.uspto.gov/sequence.html.
XX
SQ   Sequence 1032 BP; 319 A; 191 C; 281 G; 241 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:              511          Length:        1032
```

Score:	39.00	Matches:	7
Percent Similarity:	87.50%	Conservative:	0
Best Local Similarity:	87.50%	Mismatches:	1
Query Match:	81.25%	Indels:	0
DB:	13	Gaps:	0

US-10-725-373-5 (1-9) x ADS45359 (1-1032)

```
Qy          1 TyrLeuSerGlyAlaCysLeuAsn 8
              ||||||   ||||||||||||||||
Db        908 TATTTGAGAGGGGCGTGTTTAAAC 931
```

RESULT 12
AAC74312
ID AAC74312 standard; cDNA; 1384 BP.
XX
AC AAC74312;
XX
DT 02-FEB-2001 (first entry)
XX
DE Human secreted protein gene 33 SEQ ID NO:43.
XX
KW Human; secreted protein; diagnosis; immunosuppressive; antiarthritic;
KW antirheumatic; antiproliferative; cytostatic; cardiant; vasotropic;
KW cerebroprotective; nootropic; neuroprotective; antibacterial; virucide;
KW fungicide; ophthalmological; vulnerary; gene therapy; angiogenesis;
KW autoimmune disease; hyperproliferative disorder; infection; skin aging;
KW wound healing; cardiovascular disorder; cerebrovascular disorder;
KW nervous system disorder; food additive; preservative; ss.
XX
OS Homo sapiens.
XX
PN WO200057903-A2.
XX
PD 05-OCT-2000.
XX
PF 22-MAR-2000; 2000WO-US007525.
XX
PR 26-MAR-1999; 99US-0126595P.
PR 22-DEC-1999; 99US-0171549P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Ruben SM, Komatsoulis G;
XX
DR WPI; 2000-594630/56.
DR P-PSDB; AAB39342.
XX
PT New nucleic acid molecules encoding 48 human secreted proteins for
PT diagnosing, preventing, treating or ameliorating medical conditions and
PT used as food additives or preservatives.
XX
PS Claim 1; Page 339-340; 395pp; English.
XX
CC The polynucleotide sequences given in AAC74280 to AAC74327 encode the
CC human secreted proteins given in AAB39310 to AAB39357. AAB39358 to
CC AAB39400 represent human secreted polypeptide sequences and proteins
CC homologous to them, which are given in the exemplification of the present
CC invention. Human secreted proteins have activities based on the tissues
CC and cells the genes are expressed in. Examples of activities include:
CC antiarthritic; immunosuppressive; antirheumatic; antiproliferative;
CC cytostatic; cardiant; vasotropic; cerebroprotective; nootropic;
CC neuroprotective; antibacterial; virucide; fungicide; ophthalmological;
CC and vulnerary. The polynucleotides and polypeptides can be used to
CC prevent, treat or ameliorate a medical condition in e.g. humans, mice,
CC rabbits, goats, horses, cats, dogs, chickens or sheep. They can also used
CC in diagnosing a pathological condition or susceptibility to a
CC pathological condition. Disorders which are diagnosed or treated include
CC autoimmune diseases, hyperproliferative disorders, cardiovascular
CC disorders, cerebrovascular disorders, angiogenesis, nervous system
CC disorders, infections caused by bacteria, viruses and fungi and ocular
CC disorders. The polypeptides can also be used to aid wound healing and
CC epithelial cell proliferation, to prevent skin aging due to sunburn, to
CC maintain organs before transplantation, for supporting cell culture of
CC primary tissues, to regenerate tissues and in chemotaxis. The
CC polypeptides can also be used as a food additive or preservative to
CC increase or decrease storage capabilities, fat content, lipid, protein,
CC carbohydrate, vitamins, minerals, cofactors and other nutritional
CC components. AAC74271 to AAC74279 and AAB39309 represent sequences used in
CC the exemplification of the present invention
XX
SQ Sequence 1384 BP; 443 A; 247 C; 254 G; 440 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	717	Length:	1384
Score:	39.00	Matches:	7
Percent Similarity:	87.50%	Conservative:	0
Best Local Similarity:	87.50%	Mismatches:	1
Query Match:	81.25%	Indels:	0
DB:	3	Gaps:	0

US-10-725-373-5 (1-9) x AAC74312 (1-1384)

```
Qy          1 TyrLeuSerGlyAlaCysLeuAsn 8
              ||||||||||||   |||||||||
Db         23 TATTTAAGTGGAAAGTGCTTGAAC 46
```

RESULT 13
AAH53985
ID AAH53985 standard; DNA; 2950 BP.
XX
AC AAH53985;
XX
DT 03-SEP-2001 (first entry)
XX
DE S. epidermidis genomic polynucleotide sequence SEQ ID NO:3349.
XX
KW Staphylococcus epidermidis SR1 strain; infection; diagnosis; vaccination;
KW endocarditis; ds.
XX
OS Staphylococcus epidermidis.
XX
PN WO200134809-A2.
XX
PD 17-MAY-2001.
XX
PF 09-NOV-2000; 2000WO-US030782.
XX
PR 09-NOV-1999; 99US-0164258P.
XX
PA (GLAX) GLAXO GROUP LTD.
XX
PI Kimmerly WJ;
XX
DR WPI; 2001-316495/33.
XX
PT Nucleic acids encoding polypeptides from Staphylococcus epidermidis,
PT useful for vaccinating against infections, e.g. endocarditis.
XX
PS Claim 8; Page 896-897; 2188pp; English.
XX
CC AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides
CC (II), given in AAG81454 to AAG83120, from Staphylococcus epidermidis. (I)
CC and (II) can have antibacterial activity and therefore can be used in
CC vaccination. The nucleic acids (I) may be used to produce the S.
CC epidermidis polypeptides (II) via the production of vectors containing
CC them which are used to produce hosts cells which express the
CC polypeptides. The polypeptides (II) (and/or nucleic acids) may then be
CC used to vaccinate subjects and to raise antibodies against the bacteria.
CC The polypeptides may also be used to assay for other inhibitors of their
CC activity and therefore identify compounds that may be used for the
CC treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to
CC AAH55090 represent specifically claimed S. epidermidis genomic DNA
CC polynucleotide sequences from the present invention. AAH55091 to AAH55098
CC represent oligonucleotide sequences and primers which are used in the
CC exemplification of the present invention. N.B. The present invention

CC specifically claims all the polynucleotide sequences given in the
CC sequence listing of the present specification, however the sequence
CC listing only goes up to SEQ ID NO:4454 so even though sequences are given
CC in the disclosure for SEQ ID NO:4465 to 4472, no sequences are present
CC for SEQ ID NO:4455 to 4464
XX
SQ Sequence 2950 BP; 887 A; 541 C; 388 G; 1134 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	1.71e+03	Length:	2950
Score:	39.00	Matches:	6
Percent Similarity:	88.89%	Conservative:	2
Best Local Similarity:	66.67%	Mismatches:	1
Query Match:	81.25%	Indels:	0
DB:	4	Gaps:	0

US-10-725-373-5 (1-9) x AAH53985 (1-2950)

```
Qy          1 TyrLeuSerGlyAlaCysLeuAsnLeu 9
              |||   ||||||:::|||:::||||||
Db       1477 TATTTCAGTGGCTCGTGCATTAATTTG 1503
```

RESULT 14
AAH54665
ID AAH54665 standard; DNA; 3760 BP.
XX
AC AAH54665;
XX
DT 03-SEP-2001 (first entry)
XX
DE S. epidermidis genomic polynucleotide sequence SEQ ID NO:4029.
XX
KW Staphylococcus epidermidis SR1 strain; infection; diagnosis; vaccination;
KW endocarditis; ds.
XX
OS Staphylococcus epidermidis.
XX
PN WO200134809-A2.
XX
PD 17-MAY-2001.
XX
PF 09-NOV-2000; 2000WO-US030782.
XX
PR 09-NOV-1999; 99US-0164258P.
XX
PA (GLAX) GLAXO GROUP LTD.
XX
PI Kimmerly WJ;
XX
DR WPI; 2001-316495/33.
XX
PT Nucleic acids encoding polypeptides from Staphylococcus epidermidis,
PT useful for vaccinating against infections, e.g. endocarditis.
XX
PS Claim 8; Page 1708-1709; 2188pp; English.
XX
CC AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides
CC (II), given in AAG81454 to AAG83120, from Staphylococcus epidermidis. (I)
CC and (II) can have antibacterial activity and therefore can be used in
CC vaccination. The nucleic acids (I) may be used to produce the S.
CC epidermidis polypeptides (II) via the production of vectors containing
CC them which are used to produce hosts cells which express the
CC polypeptides. The polypeptides (II) (and/or nucleic acids) may then be
CC used to vaccinate subjects and to raise antibodies against the bacteria.
CC The polypeptides may also be used to assay for other inhibitors of their
CC activity and therefore identify compounds that may be used for the
CC treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to
CC AAH55090 represent specifically claimed S. epidermidis genomic DNA
CC polynucleotide sequences from the present invention. AAH55091 to AAH55098
CC represent oligonucleotide sequences and primers which are used in the
CC exemplification of the present invention. N.B. The present invention
CC specifically claims all the polynucleotide sequences given in the
CC sequence listing of the present specification, however the sequence
CC listing only goes up to SEQ ID NO:4454 so even though sequences are given
CC in the disclosure for SEQ ID NO:4465 to 4472, no sequences are present
CC for SEQ ID NO:4455 to 4464
XX
SQ Sequence 3760 BP; 1106 A; 668 C; 505 G; 1481 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	2.26e+03	Length:	3760
Score:	39.00	Matches:	6
Percent Similarity:	88.89%	Conservative:	2
Best Local Similarity:	66.67%	Mismatches:	1
Query Match:	81.25%	Indels:	0
DB:	4	Gaps:	0

US-10-725-373-5 (1-9) x AAH54665 (1-3760)

```
Qy          1 TyrLeuSerGlyAlaCysLeuAsnLeu 9
              |||   ||||||:::|||:::||||||
Db        296 TATTTCAGTGGCTCGTGCATTAATTTG 322
```

RESULT 15
ADQ97382/c
ID ADQ97382 standard; DNA; 41202 BP.
XX
AC ADQ97382;
XX
DT 07-OCT-2004 (first entry)
XX
DE Mouse cancer associated sequence MD08-041, SEQ ID 359.
XX
KW Cytostatic; Gene Therapy; cancer; leukemia; lymphoma; Mouse; ds.
XX
OS Mus musculus.
XX
PN WO2004060304-A2.
XX
PD 22-JUL-2004.
XX
PF 22-DEC-2003; 2003WO-US041389.
XX
PR 27-DEC-2002; 2002US-00330773.
XX
PA (SAGR-) SAGRES DISCOVERY INC.
XX
PI Morris DW, Malandro MS;
XX
DR WPI; 2004-543781/52.
XX
PT New isolated cancer associated nucleic acids comprising at least 10
PT contiguous nucleotides, useful for diagnosing, preventing and/or treating
PT cancers such as leukemia and lymphoma.
XX
PS Claim 1; SEQ ID NO 359; 199pp; English.
XX
CC The present invention relates to cancer associated sequences (ADQ97025-
CC ADQ98004). The sequences are useful for the diagnosis, prevention and/or
CC treatment of cancer, such as leukemia and lymphoma. Note: The sequence
CC data for this patent did not form part of the printed specification, but
CC was obtained in electronic formate directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 41202 BP; 9227 A; 11011 C; 11099 G; 9687 T; 0 U; 178 Other;

Alignment Scores:

Pred. No.:	3.55e+04	Length:	41202
Score:	39.00	Matches:	7
Percent Similarity:	88.89%	Conservative:	1
Best Local Similarity:	77.78%	Mismatches:	1
Query Match:	81.25%	Indels:	0
DB:	12	Gaps:	0

US-10-725-373-5 (1-9) x ADQ97382 (1-41202)

```
Qy          1 TyrLeuSerGlyAlaCysLeuAsnLeu 9
              ||||||:::|||||||||||   |||
Db      18116 TATCTCACAGGGGCCTGCCTAGGACTC 18090
```

Search completed: May 17, 2005, 17:45:21
Job time : 339.5 secs

GenCore version 5.1.6

OM protein - nucleic search, using frame_plus_p2n model

Run on: May 17, 2005, 16:32:09 ; Search time 1297.5 Seconds
(without alignments)
336.106 Million cell updates/sec

Title: US-10-725-373-4
Perfect score: 45
Sequence: 1 YLSGANINL 9

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/US10725373/runat_17052005_071020_16185/app_query.fasta_1.796
-DB=GenEmbl -QFMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10725373_@CGN_1_1_6148_@runat_17052005_071020_16185 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	45	100.0	27	6	AR560607	AR560607 Sequence
	2	45	100.0	27	6	BD131678	BD131678 Carcinoem
	3	43	95.6	27	6	AR560604	AR560604 Sequence
	4	43	95.6	27	6	BD131675	BD131675 Carcinoem
c	5	43	95.6	80	6	A87744	A87744 Sequence 38
c	6	43	95.6	80	6	BD195831	BD195831 Method fo
	7	43	95.6	155	6	AX193221	AX193221 Sequence
	8	43	95.6	256	6	AX260775	AX260775 Sequence
c	9	43	95.6	407	6	AX260391	AX260391 Sequence
	10	43	95.6	409	6	AR273297	AR273297 Sequence
	11	43	95.6	409	6	AR273585	AR273585 Sequence
c	12	43	95.6	409	6	AR273719	AR273719 Sequence
	13	43	95.6	409	6	AR276878	AR276878 Sequence
	14	43	95.6	409	6	AR277166	AR277166 Sequence
c	15	43	95.6	409	6	AR277300	AR277300 Sequence
	16	43	95.6	409	6	AR407153	AR407153 Sequence
	17	43	95.6	409	6	AR407441	AR407441 Sequence
c	18	43	95.6	409	6	AR407575	AR407575 Sequence
	19	43	95.6	409	6	AR441003	AR441003 Sequence
	20	43	95.6	409	6	AR441291	AR441291 Sequence
c	21	43	95.6	409	6	AR441425	AR441425 Sequence
	22	43	95.6	409	6	AR543814	AR543814 Sequence
	23	43	95.6	409	6	AR544102	AR544102 Sequence
c	24	43	95.6	409	6	AR544236	AR544236 Sequence
c	25	43	95.6	409	6	AX260450	AX260450 Sequence
c	26	43	95.6	409	6	AX351472	AX351472 Sequence
	27	43	95.6	409	6	AX368330	AX368330 Sequence
	28	43	95.6	409	6	AX368618	AX368618 Sequence
c	29	43	95.6	409	6	AX368752	AX368752 Sequence
	30	43	95.6	409	6	AX396625	AX396625 Sequence
	31	43	95.6	409	6	AX397282	AX397282 Sequence
c	32	43	95.6	410	6	AX260455	AX260455 Sequence
	33	43	95.6	410	6	AX260459	AX260459 Sequence
	34	43	95.6	412	6	AR166812	AR166812 Sequence
	35	43	95.6	412	6	BD265007	BD265007 Compounds
	36	43	95.6	412	6	AR400993	AR400993 Sequence
	37	43	95.6	412	6	AX192438	AX192438 Sequence
	38	43	95.6	413	6	AX260652	AX260652 Sequence
	39	43	95.6	415	6	AR272706	AR272706 Sequence
	40	43	95.6	415	6	AR276287	AR276287 Sequence
	41	43	95.6	415	6	AR406562	AR406562 Sequence
	42	43	95.6	415	6	AR440412	AR440412 Sequence
	43	43	95.6	415	6	AR472570	AR472570 Sequence
	44	43	95.6	415	6	AR543223	AR543223 Sequence
	45	43	95.6	415	6	AX062808	AX062808 Sequence

ALIGNMENTS

RESULT 1
AR560607
LOCUS AR560607 27 bp DNA linear PAT 08-OCT-2004
DEFINITION Sequence 11 from patent US 6756038.
ACCESSION AR560607
VERSION AR560607.1 GI:53972928
KEYWORDS .
SOURCE Unknown.
ORGANISM Unknown.
Unclassified.
REFERENCE 1 (bases 1 to 27)
AUTHORS Schlom,J., Barzaga,E. and Zaremba,S.
TITLE Agonist and antagonist peptides of carcinoembryonic antigen (CEA)
JOURNAL Patent: US 6756038-A 11 29-JUN-2004;
FEATURES Location/Qualifiers
source 1. .27
/organism="unknown"
/mol_type="genomic DNA"
ORIGIN

Alignment Scores:

Pred. No.:	0.00992	Length:	27
Score:	45.00	Matches:	9
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	6	Gaps:	0

```
US-10-725-373-4 (1-9) x AR560607 (1-27)

Qy          1 TyrLeuSerGlyAlaAsnIleAsnLeu 9
              ||||||||||||||||||||||||||
Db          1 TACCTTTCGGGAGCGAACATCAACCTC 27

RESULT 2
BD131678
LOCUS       BD131678                  27 bp    DNA     linear   PAT 18-SEP-2002
DEFINITION  Carcinoembryonic antigen (CEA) agonist and antagonist peptides.
ACCESSION   BD131678
VERSION     BD131678.1  GI:23226623
KEYWORDS    JP 2002500002-A/4.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 27)
  AUTHORS   Schlom,J., Barzaga,E. and Zaremba,S.
  TITLE     Carcinoembryonic antigen (CEA) agonist and antagonist peptides
  JOURNAL   Patent: JP 2002500002-A 4 08-JAN-2002;
            THE UNITED STATES OF AMERICA
COMMENT     OS   Homo sapiens (human)
            PN   JP 2002500002-A/4
            PD   08-JAN-2002
            PF   22-SEP-1998 JP 2000516030
            PR   10-OCT-1997 US    60/061589
            PI   JEFFREY SCHLOM,ELENE BARZAGA,SAM ZAREMBA
            PC   C12N15/09,A61K38/00,A61K45/00,A61K48/00,A61P35/00,A61P37/02,
            PC   A61P43/00,
            PC   C07K14/705,C07K16/28,C12N5/10,C12N15/00,A61K37/02,C12N5/00 CC
             Carcinoembryonic antigen (CEA) agonist and antagonist peptides FH
             Key             Location/Qualifiers
            FT   source          1. .27
            FT                   /organism='Homo sapiens (human)'.
FEATURES             Location/Qualifiers
     source          1. .27
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
ORIGIN

Alignment Scores:
Pred. No.:              0.00992         Length:         27
Score:                  45.00           Matches:        9
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     6               Gaps:           0

US-10-725-373-4 (1-9) x BD131678 (1-27)

Qy          1 TyrLeuSerGlyAlaAsnIleAsnLeu 9
              ||||||||||||||||||||||||||
Db          1 TACCTTTCGGGAGCGAACATCAACCTC 27

RESULT 3
AR560604
LOCUS       AR560604                  27 bp    DNA     linear   PAT 08-OCT-2004
DEFINITION  Sequence 6 from patent US 6756038.
ACCESSION   AR560604
VERSION     AR560604.1  GI:53972925
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 27)
  AUTHORS   Schlom,J., Barzaga,E. and Zaremba,S.
  TITLE     Agonist and antagonist peptides of carcinoembryonic antigen (CEA)
  JOURNAL   Patent: US 6756038-A 6 29-JUN-2004;
FEATURES             Location/Qualifiers
     source          1. .27
                     /organism="unknown"
                     /mol_type="genomic DNA"
ORIGIN

Alignment Scores:
Pred. No.:              0.0311          Length:         27
Score:                  43.00           Matches:        8
Percent Similarity:     100.00%         Conservative:   1
Best Local Similarity:  88.89%          Mismatches:     0
Query Match:            95.56%          Indels:         0
DB:                     6               Gaps:           0

US-10-725-373-4 (1-9) x AR560604 (1-27)

Qy          1 TyrLeuSerGlyAlaAsnIleAsnLeu 9
              ||||||||||||||||:::||||||
Db          1 TACCTTTCGGGAGCGAACCTCAACCTC 27

RESULT 4
BD131675
LOCUS       BD131675                  27 bp    DNA     linear   PAT 18-SEP-2002
DEFINITION  Carcinoembryonic antigen (CEA) agonist and antagonist peptides.
ACCESSION   BD131675
VERSION     BD131675.1  GI:23226620
KEYWORDS    JP 2002500002-A/1.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 27)
  AUTHORS   Schlom,J., Barzaga,E. and Zaremba,S.
  TITLE     Carcinoembryonic antigen (CEA) agonist and antagonist peptides
  JOURNAL   Patent: JP 2002500002-A 1 08-JAN-2002;
            THE UNITED STATES OF AMERICA
COMMENT     OS   Homo sapiens (human)
            PN   JP 2002500002-A/1
            PD   08-JAN-2002
            PF   22-SEP-1998 JP 2000516030
            PR   10-OCT-1997 US    60/061589
            PI   JEFFREY SCHLOM,ELENE BARZAGA,SAM ZAREMBA
            PC   C12N15/09,A61K38/00,A61K45/00,A61K48/00,A61P35/00,A61P37/02,
            PC   A61P43/00,
            PC   C07K14/705,C07K16/28,C12N5/10,C12N15/00,A61K37/02,C12N5/00 CC
             Carcinoembryonic antigen (CEA) agonist and antagonist peptides FH
             Key             Location/Qualifiers
            FT   source          1. .27
            FT                   /organism='Homo sapiens (human)'.
FEATURES             Location/Qualifiers
     source          1. .27
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
ORIGIN

Alignment Scores:
Pred. No.:              0.0311          Length:         27
Score:                  43.00           Matches:        8
Percent Similarity:     100.00%         Conservative:   1
Best Local Similarity:  88.89%          Mismatches:     0
Query Match:            95.56%          Indels:         0
DB:                     6               Gaps:           0

US-10-725-373-4 (1-9) x BD131675 (1-27)

Qy          1 TyrLeuSerGlyAlaAsnIleAsnLeu 9
              ||||||||||||||||:::||||||
Db          1 TACCTTTCGGGAGCGAACCTCAACCTC 27

RESULT 5
A87744/c
LOCUS       A87744                    80 bp    DNA     linear   PAT 22-JAN-2000
DEFINITION  Sequence 38 from Patent WO9833523.
ACCESSION   A87744
VERSION     A87744.1  GI:6736346
```

```
KEYWORDS    .
SOURCE      unidentified
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 80)
  AUTHORS   Carr,F.J. and Carter,G.
  TITLE     VACCINATION METHODS AND MOLECULES
  JOURNAL   Patent: WO 9833523-A 38 06-AUG-1998;
            BIOVATION LIMITED (GB); CARR FRANK JOSEPH (GB)
FEATURES             Location/Qualifiers
     source          1. .80
                     /organism="unidentified"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32644"
ORIGIN

Alignment Scores:
Pred. No.:              0.108          Length:         80
Score:                  43.00          Matches:        8
Percent Similarity:     100.00%        Conservative:   1
Best Local Similarity:  88.89%         Mismatches:     0
Query Match:            95.56%         Indels:         0
DB:                     6              Gaps:           0

US-10-725-373-4 (1-9) x A87744 (1-80)

Qy          1 TyrLeuSerGlyAlaAsnIleAsnLeu 9
              ||||||||||||||||||:::||||||
Db         68 TACCTGTCCGGCGCCAACCTGAACCTG 42

RESULT 6
BD195831/c
LOCUS       BD195831                80 bp    DNA     linear   PAT 17-JUL-2003
DEFINITION  Method for the production of non-immunogenic proteins.
ACCESSION   BD195831
VERSION     BD195831.1  GI:33005601
KEYWORDS    JP 2002512624-A/102.
SOURCE      unidentified
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 80)
  AUTHORS   Carr,F.J., Adair,F.S., Hamilton,A.A. and Carter,G.
  TITLE     Method for the production of non-immunogenic proteins
  JOURNAL   Patent: JP 2002512624-A 102 23-APR-2002;
            BIOVATION LTD
COMMENT     OS   Unidentified
            PN   JP 2002512624-A/102
            PD   23-APR-2002
            PF   21-MAY-1998 JP 1998550129
            PR   21-MAY-1997 GB    9710480.6,31-JUL-1997 GB    9716197.0 PR
            28-NOV-1997 GB    9725270.4,02-DEC-1997 US    60/067235 PR
            14-APR-1998 GB    9807751.4
            PI   FRANCIS JOSEPH CARR,FIONA SUZANNE ADAIR,ANITA ANNE HAMILTON,
            PI   GRAHAM CARTER
            PC   C07K16/46,C07K14/315,G01N33/563,A61K39/395
            CC   Strandedness: Single;
            CC   Topology: Linear;
            CC   Method for the production of non-immunogenic proteins FH   Key
                    Location/Qualifiers
            FT   source          1. .80
            FT                   /organism='Unidentified'.
FEATURES             Location/Qualifiers
     source          1. .80
                     /organism="unidentified"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:32644"
ORIGIN

Alignment Scores:
Pred. No.:              0.108          Length:         80
Score:                  43.00          Matches:        8
Percent Similarity:     100.00%        Conservative:   1
Best Local Similarity:  88.89%         Mismatches:     0
Query Match:            95.56%         Indels:         0
DB:                     6              Gaps:           0

US-10-725-373-4 (1-9) x BD195831 (1-80)

Qy          1 TyrLeuSerGlyAlaAsnIleAsnLeu 9
              ||||||||||||||||||:::||||||
Db         68 TACCTGTCCGGCGCCAACCTGAACCTG 42

RESULT 7
AX193221
LOCUS       AX193221                155 bp    DNA     linear   PAT 15-AUG-2001
DEFINITION  Sequence 788 from Patent WO0149716.
ACCESSION   AX193221
VERSION     AX193221.1  GI:15211172
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Xu,J., Lodes,M.J., Secrist,H., Benson,D.R., Meagher,M.J.,
            Stolk,J.A., King,G.E., Wang,T. and Jiang,Y.
  TITLE     Compounds for immunotherapy and diagnosis of colon cancer and
            methods for their use
  JOURNAL   Patent: WO 0149716-A 788 12-JUL-2001;
            CORIXA CORPORATION (US)
FEATURES             Location/Qualifiers
     source          1. .155
                     /organism="Homo sapiens"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:9606"
ORIGIN

Alignment Scores:
Pred. No.:              0.23           Length:         155
Score:                  43.00          Matches:        8
Percent Similarity:     100.00%        Conservative:   1
Best Local Similarity:  88.89%         Mismatches:     0
Query Match:            95.56%         Indels:         0
DB:                     6              Gaps:           0

US-10-725-373-4 (1-9) x AX193221 (1-155)

Qy          1 TyrLeuSerGlyAlaAsnIleAsnLeu 9
              ||||||||||||||||||:::||||||
Db        119 TACCTTTCNGGAGCGAACCTCAACCTC 145

RESULT 8
AX260775
LOCUS       AX260775                256 bp    DNA     linear   PAT 26-OCT-2001
DEFINITION  Sequence 426 from Patent WO0173027.
ACCESSION   AX260775
VERSION     AX260775.1  GI:16509742
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Meagher,M.J., Xu,J. and King,G.E.
  TITLE     Compositions and methods for therapy and diagnosis of colon cancer
  JOURNAL   Patent: WO 0173027-A 426 04-OCT-2001;
            CORIXA CORPORATION (US)
FEATURES             Location/Qualifiers
     source          1. .256
                     /organism="Homo sapiens"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:9606"
ORIGIN

Alignment Scores:
Pred. No.:              0.41           Length:         256
```

```
Score:                  43.00          Matches:       8
Percent Similarity:     100.00%        Conservative:  1
Best Local Similarity:  88.89%         Mismatches:    0
Query Match:            95.56%         Indels:        0
DB:                     6              Gaps:          0

US-10-725-373-4 (1-9) x AX260775 (1-256)

Qy          1 TyrLeuSerGlyAlaAsnIleAsnLeu 9
              ||||||||||||||||||:::||||||
Db         15 TACCTTTCGGGAGCGAACCTCAACCTC 41

RESULT 9
AX260391/c
LOCUS       AX260391                407 bp    DNA     linear   PAT 26-OCT-2001
DEFINITION  Sequence 42 from Patent WO0173027.
ACCESSION   AX260391
VERSION     AX260391.1  GI:16509350
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Meagher,M.J., Xu,J. and King,G.E.
  TITLE     Compositions and methods for therapy and diagnosis of colon cancer
  JOURNAL   Patent: WO 0173027-A 42 04-OCT-2001;
            CORIXA CORPORATION (US)
FEATURES             Location/Qualifiers
     source          1. .407
                     /organism="Homo sapiens"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:9606"
ORIGIN

Alignment Scores:
Pred. No.:              0.697          Length:        407
Score:                  43.00          Matches:       8
Percent Similarity:     100.00%        Conservative:  1
Best Local Similarity:  88.89%         Mismatches:    0
Query Match:            95.56%         Indels:        0
DB:                     6              Gaps:          0

US-10-725-373-4 (1-9) x AX260391 (1-407)

Qy          1 TyrLeuSerGlyAlaAsnIleAsnLeu 9
              ||||||||||||||||||:::||||||
Db        241 TACCTTTCGGGAGCGAACCTCAACCTC 215

RESULT 10
AR273297
LOCUS       AR273297                409 bp    DNA     linear   PAT 10-APR-2003
DEFINITION  Sequence 1040 from patent US 6504010.
ACCESSION   AR273297
VERSION     AR273297.1  GI:29705182
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 409)
  AUTHORS   Wang,T., Bangur,C.S., Lodes,M.J., Fanger,G.R., Vedvick,T.S.,
            Carter,D., Retter,M.W., Mannion,J. and Fan,L.
  TITLE     Compositions and methods for the therapy and diagnosis of lung
            cancer
  JOURNAL   Patent: US 6504010-A 1040 07-JAN-2003;
FEATURES             Location/Qualifiers
     source          1. .409
                     /organism="unknown"
                     /mol_type="genomic DNA"
ORIGIN

Alignment Scores:
Pred. No.:              0.701          Length:        409
```

```
Score:                  43.00          Matches:       8
Percent Similarity:     100.00%        Conservative:  1
Best Local Similarity:  88.89%         Mismatches:    0
Query Match:            95.56%         Indels:        0
DB:                     6              Gaps:          0

US-10-725-373-4 (1-9) x AR273297 (1-409)

Qy          1 TyrLeuSerGlyAlaAsnIleAsnLeu 9
              ||||||||||||||||||:::||||||
Db        169 TACCTTTCGGGAGCGAACCTCAACCTC 195

RESULT 11
AR273585
LOCUS       AR273585                409 bp    DNA     linear   PAT 10-APR-2003
DEFINITION  Sequence 1328 from patent US 6504010.
ACCESSION   AR273585
VERSION     AR273585.1  GI:29705470
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 409)
  AUTHORS   Wang,T., Bangur,C.S., Lodes,M.J., Fanger,G.R., Vedvick,T.S.,
            Carter,D., Retter,M.W., Mannion,J. and Fan,L.
  TITLE     Compositions and methods for the therapy and diagnosis of lung
            cancer
  JOURNAL   Patent: US 6504010-A 1328 07-JAN-2003;
FEATURES             Location/Qualifiers
     source          1. .409
                     /organism="unknown"
                     /mol_type="genomic DNA"
ORIGIN

Alignment Scores:
Pred. No.:              0.701          Length:        409
Score:                  43.00          Matches:       8
Percent Similarity:     100.00%        Conservative:  1
Best Local Similarity:  88.89%         Mismatches:    0
Query Match:            95.56%         Indels:        0
DB:                     6              Gaps:          0

US-10-725-373-4 (1-9) x AR273585 (1-409)

Qy          1 TyrLeuSerGlyAlaAsnIleAsnLeu 9
              ||||||||||||||||||:::||||||
Db        169 TACCTTTCGGGAGCGAACCTCAACCTC 195

RESULT 12
AR273719/c
LOCUS       AR273719                409 bp    DNA     linear   PAT 10-APR-2003
DEFINITION  Sequence 1462 from patent US 6504010.
ACCESSION   AR273719
VERSION     AR273719.1  GI:29705604
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 409)
  AUTHORS   Wang,T., Bangur,C.S., Lodes,M.J., Fanger,G.R., Vedvick,T.S.,
            Carter,D., Retter,M.W., Mannion,J. and Fan,L.
  TITLE     Compositions and methods for the therapy and diagnosis of lung
            cancer
  JOURNAL   Patent: US 6504010-A 1462 07-JAN-2003;
FEATURES             Location/Qualifiers
     source          1. .409
                     /organism="unknown"
                     /mol_type="genomic DNA"
ORIGIN

Alignment Scores:
Pred. No.:              0.701          Length:        409
Score:                  43.00          Matches:       8
```

```
Percent Similarity:     100.00%         Conservative:  1
Best Local Similarity:  88.89%          Mismatches:    0
Query Match:            95.56%          Indels:        0
DB:                     6               Gaps:          0

US-10-725-373-4 (1-9) x AR273719 (1-409)

Qy          1 TyrLeuSerGlyAlaAsnIleAsnLeu 9
              ||||||||||||||||||:::||||||
Db        241 TACCTTTCGGGAGCGAACCTCAACCTC 215

RESULT 13
AR276878
LOCUS       AR276878                409 bp    DNA     linear   PAT 10-APR-2003
DEFINITION  Sequence 1040 from patent US 6509448.
ACCESSION   AR276878
VERSION     AR276878.1  GI:29710525
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 409)
  AUTHORS   Wang,T., Bangur,C.S., Lodes,M.J., Fanger,G.R., Vedvick,T.S.,
            Carter,D., Retter,M.W., Mannion,J., Fan,L. and Wang,A.
  TITLE     Compositions and methods for the therapy and diagnosis of lung
            cancer
  JOURNAL   Patent: US 6509448-A 1040 21-JAN-2003;
FEATURES             Location/Qualifiers
     source          1. .409
                     /organism="unknown"
                     /mol_type="genomic DNA"
ORIGIN

Alignment Scores:
Pred. No.:              0.701           Length:        409
Score:                  43.00           Matches:       8
Percent Similarity:     100.00%         Conservative:  1
Best Local Similarity:  88.89%          Mismatches:    0
Query Match:            95.56%          Indels:        0
DB:                     6               Gaps:          0

US-10-725-373-4 (1-9) x AR276878 (1-409)

Qy          1 TyrLeuSerGlyAlaAsnIleAsnLeu 9
              ||||||||||||||||||:::||||||
Db        169 TACCTTTCGGGAGCGAACCTCAACCTC 195

RESULT 14
AR277166
LOCUS       AR277166                409 bp    DNA     linear   PAT 10-APR-2003
DEFINITION  Sequence 1328 from patent US 6509448.
ACCESSION   AR277166
VERSION     AR277166.1  GI:29710813
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 409)
  AUTHORS   Wang,T., Bangur,C.S., Lodes,M.J., Fanger,G.R., Vedvick,T.S.,
            Carter,D., Retter,M.W., Mannion,J., Fan,L. and Wang,A.
  TITLE     Compositions and methods for the therapy and diagnosis of lung
            cancer
  JOURNAL   Patent: US 6509448-A 1328 21-JAN-2003;
FEATURES             Location/Qualifiers
     source          1. .409
                     /organism="unknown"
                     /mol_type="genomic DNA"
ORIGIN

Alignment Scores:
Pred. No.:              0.701           Length:        409
Score:                  43.00           Matches:       8
Percent Similarity:     100.00%         Conservative:  1
Best Local Similarity:  88.89%          Mismatches:    0
Query Match:            95.56%          Indels:        0
DB:                     6               Gaps:          0

US-10-725-373-4 (1-9) x AR277166 (1-409)

Qy          1 TyrLeuSerGlyAlaAsnIleAsnLeu 9
              ||||||||||||||||||:::||||||
Db        169 TACCTTTCGGGAGCGAACCTCAACCTC 195

RESULT 15
AR277300/c
LOCUS       AR277300                409 bp    DNA     linear   PAT 10-APR-2003
DEFINITION  Sequence 1462 from patent US 6509448.
ACCESSION   AR277300
VERSION     AR277300.1  GI:29710947
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 409)
  AUTHORS   Wang,T., Bangur,C.S., Lodes,M.J., Fanger,G.R., Vedvick,T.S.,
            Carter,D., Retter,M.W., Mannion,J., Fan,L. and Wang,A.
  TITLE     Compositions and methods for the therapy and diagnosis of lung
            cancer
  JOURNAL   Patent: US 6509448-A 1462 21-JAN-2003;
FEATURES             Location/Qualifiers
     source          1. .409
                     /organism="unknown"
                     /mol_type="genomic DNA"
ORIGIN

Alignment Scores:
Pred. No.:              0.701           Length:        409
Score:                  43.00           Matches:       8
Percent Similarity:     100.00%         Conservative:  1
Best Local Similarity:  88.89%          Mismatches:    0
Query Match:            95.56%          Indels:        0
DB:                     6               Gaps:          0

US-10-725-373-4 (1-9) x AR277300 (1-409)

Qy          1 TyrLeuSerGlyAlaAsnIleAsnLeu 9
              ||||||||||||||||||:::||||||
Db        241 TACCTTTCGGGAGCGAACCTCAACCTC 215


Search completed: May 17, 2005, 19:12:54
Job time : 1298.5 secs
```

GenCore version 5.1.6

OM protein - nucleic search, using frame_plus_p2n model

Run on: May 17, 2005, 16:32:09 ; Search time 1297.5 Seconds
(without alignments)
336.106 Million cell updates/sec

Title: US-10-725-373-3
Perfect score: 45
Sequence: 1 YLSGADINL 9

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/US10725373/runat_17052005_071020_16185/app_query.fasta_1.796
-DB=GenEmbl -QFMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10725373_@CGN_1_1_6148_@runat_17052005_071020_16185 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	45	100.0	27	6	AR560606	AR560606 Sequence
2	45	100.0	27	6	BD131677	BD131677 Carcinoem
3	43	95.6	27	6	AR560605	AR560605 Sequence
4	43	95.6	27	6	BD131676	BD131676 Carcinoem
5	43	95.6	2106	6	AX133657	AX133657 Sequence
6	43	95.6	2106	6	AX192349	AX192349 Sequence
7	43	95.6	2106	6	AX393888	AX393888 Sequence
8	42	93.3	151138	2	AC136727	AC136727 Mus muscu
c 9	42	93.3	185385	2	AC149846	AC149846 Papio anu
10	41	91.1	8239	8	AY115485	AY115485 Zea mays
c 11	41	91.1	159231	9	AL161654	AL161654 Human DNA
c 12	41	91.1	168608	2	AL591477	AL591477 Homo sapi
13	40	88.9	27	6	AR560607	AR560607 Sequence
14	40	88.9	27	6	BD131678	BD131678 Carcinoem
c 15	40	88.9	108845	6	CQ869594	CQ869594 Sequence
c 16	40	88.9	154262	2	AC127629	AC127629 Rattus no
17	40	88.9	186854	2	AC110185	AC110185 Mus muscu
18	40	88.9	192525	2	AC109540	AC109540 Rattus no
c 19	40	88.9	196838	2	AC115196	AC115196 Rattus no
c 20	40	88.9	282895	2	AC096056	AC096056 Rattus no
21	39	86.7	206	8	AF155309	AF155309 Otospermu
c 22	39	86.7	605	8	FU419931	AJ419931 Rhizocton
c 23	39	86.7	611	8	FU419929	AJ419929 Rhizocton
c 24	39	86.7	613	8	CSP427400	AJ427400 Ceratobas
c 25	39	86.7	659	8	RSO318430	AJ318430 Rhizocton
c 26	39	86.7	659	8	RSO318431	AJ318431 Rhizocton
c 27	39	86.7	659	8	RSP318421	AJ318421 Rhizocton
c 28	39	86.7	659	8	RSP318422	AJ318422 Rhizocton
c 29	39	86.7	659	8	RSP318423	AJ318423 Rhizocton
c 30	39	86.7	659	8	RSP318424	AJ318424 Rhizocton
c 31	39	86.7	659	8	RSP318425	AJ318425 Rhizocton
c 32	39	86.7	659	8	RSP318428	AJ318428 Rhizocton
c 33	39	86.7	1536	8	AF354095	AF354095 Ceratobas
c 34	39	86.7	2000	6	AX510026	AX510026 Sequence
35	39	86.7	100310	2	F17A13	AL096692 Arabidops
36	39	86.7	147640	5	BX323035	BX323035 Zebrafish
c 37	39	86.7	148904	5	BX088527	BX088527 Zebrafish
38	39	86.7	155910	9	AC146104	AC146104 Pan trogl
c 39	39	86.7	157511	9	AC074348	AC074348 Homo sapi
40	39	86.7	166541	2	AC144873	AC144873 Pongo pyg
c 41	39	86.7	172980	10	AC126671	AC126671 Mus muscu
c 42	39	86.7	179521	5	AL929374	AL929374 Zebrafish
43	39	86.7	198429	8	ATCHRIV70	AL161574 Arabidops
c 44	39	86.7	251303	2	AC114862	AC114862 Rattus no
c 45	39	86.7	255451	2	AC110713	AC110713 Rattus no

ALIGNMENTS

RESULT 1
AR560606
LOCUS AR560606 27 bp DNA linear PAT 08-OCT-2004
DEFINITION Sequence 8 from patent US 6756038.
ACCESSION AR560606
VERSION AR560606.1 GI:53972927
KEYWORDS .
SOURCE Unknown.
ORGANISM Unknown.
Unclassified.
REFERENCE 1 (bases 1 to 27)
AUTHORS Schlom,J., Barzaga,E. and Zaremba,S.
TITLE Agonist and antagonist peptides of carcinoembryonic antigen (CEA)
JOURNAL Patent: US 6756038-A 8 29-JUN-2004;
FEATURES Location/Qualifiers
source 1. .27
/organism="unknown"
/mol_type="genomic DNA"
ORIGIN

Alignment Scores:

Pred. No.:	0.00346	Length:	27
Score:	45.00	Matches:	9
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	6	Gaps:	0

```
US-10-725-373-3 (1-9) x AR560606 (1-27)

Qy          1 TyrLeuSerGlyAlaAspIleAsnLeu 9
              |||||||||||||||||||||||||||
Db          1 TACCTTTCGGGAGCGGACATCAACCTC 27

RESULT 2
BD131677
LOCUS       BD131677                  27 bp    DNA     linear   PAT 18-SEP-2002
DEFINITION  Carcinoembryonic antigen (CEA) agonist and antagonist peptides.
ACCESSION   BD131677
VERSION     BD131677.1  GI:23226622
KEYWORDS    JP 2002500002-A/3.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 27)
  AUTHORS   Schlom,J., Barzaga,E. and Zaremba,S.
  TITLE     Carcinoembryonic antigen (CEA) agonist and antagonist peptides
  JOURNAL   Patent: JP 2002500002-A 3 08-JAN-2002;
            THE UNITED STATES OF AMERICA
COMMENT     OS   Homo sapiens (human)
            PN   JP 2002500002-A/3
            PD   08-JAN-2002
            PF   22-SEP-1998 JP 2000516030
            PR   10-OCT-1997 US    60/061589
            PI   JEFFREY SCHLOM,ELENE BARZAGA,SAM ZAREMBA
            PC   C12N15/09,A61K38/00,A61K45/00,A61K48/00,A61P35/00,A61P37/02,
            PC   A61P43/00,
            PC   C07K14/705,C07K16/28,C12N5/10,C12N15/00,A61K37/02,C12N5/00 CC
             Carcinoembryonic antigen (CEA) agonist and antagonist peptides FH
              Key             Location/Qualifiers
            FT   source          1. .27
            FT                   /organism='Homo sapiens (human)'.
FEATURES             Location/Qualifiers
     source          1. .27
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
ORIGIN

Alignment Scores:
Pred. No.:              0.00346         Length:         27
Score:                  45.00           Matches:        9
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     6               Gaps:           0

US-10-725-373-3 (1-9) x BD131677 (1-27)

Qy          1 TyrLeuSerGlyAlaAspIleAsnLeu 9
              |||||||||||||||||||||||||||
Db          1 TACCTTTCGGGAGCGGACATCAACCTC 27

RESULT 3
AR560605
LOCUS       AR560605                  27 bp    DNA     linear   PAT 08-OCT-2004
DEFINITION  Sequence 7 from patent US 6756038.
ACCESSION   AR560605
VERSION     AR560605.1  GI:53972926
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 27)
  AUTHORS   Schlom,J., Barzaga,E. and Zaremba,S.
  TITLE     Agonist and antagonist peptides of carcinoembryonic antigen (CEA)
  JOURNAL   Patent: US 6756038-A 7 29-JUN-2004;
FEATURES             Location/Qualifiers
     source          1. .27
                     /organism="unknown"
                     /mol_type="genomic DNA"
ORIGIN

Alignment Scores:
Pred. No.:              0.0114          Length:         27
Score:                  43.00           Matches:        8
Percent Similarity:     100.00%         Conservative:   1
Best Local Similarity:  88.89%          Mismatches:     0
Query Match:            95.56%          Indels:         0
DB:                     6               Gaps:           0

US-10-725-373-3 (1-9) x AR560605 (1-27)

Qy          1 TyrLeuSerGlyAlaAspIleAsnLeu 9
              |||||||||||||||||:::||||||
Db          1 TACCTTTCGGGAGCGGACCTCAACCTC 27

RESULT 4
BD131676
LOCUS       BD131676                  27 bp    DNA     linear   PAT 18-SEP-2002
DEFINITION  Carcinoembryonic antigen (CEA) agonist and antagonist peptides.
ACCESSION   BD131676
VERSION     BD131676.1  GI:23226621
KEYWORDS    JP 2002500002-A/2.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 27)
  AUTHORS   Schlom,J., Barzaga,E. and Zaremba,S.
  TITLE     Carcinoembryonic antigen (CEA) agonist and antagonist peptides
  JOURNAL   Patent: JP 2002500002-A 2 08-JAN-2002;
            THE UNITED STATES OF AMERICA
COMMENT     OS   Homo sapiens (human)
            PN   JP 2002500002-A/2
            PD   08-JAN-2002
            PF   22-SEP-1998 JP 2000516030
            PR   10-OCT-1997 US    60/061589
            PI   JEFFREY SCHLOM,ELENE BARZAGA,SAM ZAREMBA
            PC   C12N15/09,A61K38/00,A61K45/00,A61K48/00,A61P35/00,A61P37/02,
            PC   A61P43/00,
            PC   C07K14/705,C07K16/28,C12N5/10,C12N15/00,A61K37/02,C12N5/00 CC
             Carcinoembryonic antigen (CEA) agonist and antagonist peptides FH
              Key             Location/Qualifiers
            FT   source          1. .27
            FT                   /organism='Homo sapiens (human)'.
FEATURES             Location/Qualifiers
     source          1. .27
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
ORIGIN

Alignment Scores:
Pred. No.:              0.0114          Length:         27
Score:                  43.00           Matches:        8
Percent Similarity:     100.00%         Conservative:   1
Best Local Similarity:  88.89%          Mismatches:     0
Query Match:            95.56%          Indels:         0
DB:                     6               Gaps:           0

US-10-725-373-3 (1-9) x BD131676 (1-27)

Qy          1 TyrLeuSerGlyAlaAspIleAsnLeu 9
              |||||||||||||||||:::||||||
Db          1 TACCTTTCGGGAGCGGACCTCAACCTC 27

RESULT 5
AX133657
LOCUS       AX133657                2106 bp    DNA     linear   PAT 15-MAY-2001
DEFINITION  Sequence 111 from Patent WO0130382.
ACCESSION   AX133657
VERSION     AX133657.1  GI:14139699
```

```
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1
  AUTHORS   Berinstein,N., Tartaglia,J., Moingeon,P. and Barber,B.
  TITLE     Method of inducing and/or enhancing an immune response to tumor
            antigens
  JOURNAL   Patent: WO 0130382-A 111 03-MAY-2001;
            Aventis Pasteur Limited (CA)
FEATURES             Location/Qualifiers
     source          1. .2106
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
     CDS             1. .2106
                     /note="unnamed protein product; modified CEA"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAC38955.1"
                     /db_xref="GI:14139700"
                     /translation="MESPSAPPHRWCIPWQRLLLTASLLTFWNPPTTAKLTIESTPFN
                     VAEGKEVLLLVHNLPQHLFGYSWYKGERVDGNRQIIGYVIGTQQATPGPAYSGREIIY
                     PNASLLIQNIIQNDTGFYTLHVIKSDLVNEEATGQFRVYPELPKPSISSNNSKPVEDK
                     DAVAFTCEPETQDATYLWWVNNQSLPVSPRLQLSNGNRTLTLFNVTRNDTASYKCETQ
                     NPVSARRSDSVILNVLYGPDAPTISPLNTSYRSGENLNLSCHAASNPPAQYSWFVNGT
                     FQQSTQELFIPNITVNNSGSYTCQAHNSDTGLNRTTVTTITVYEPPKPFITSNNSNPV
                     EDEDAVALTCEPEIQNTTYLWWVNNQSLPVSPRLQLSNDNRTLTLLSVTRNDVGPYEC
                     GIQNELSVDHSDPVILNVLYGPDDPTISPSYTYYRPGVNLSLSCHAASNPPAQYSWLI
                     DGNIQQHTQELFISNITEKNSGLYTCQANNSASGHSRTTVKTITVSAELPKPSISSNN
                     SKPVEDKDAVAFTCEPEAQNTTYLWWVNGQSLPVSPRLQLSNGNRTLTLFNVTRNDAR
                     AYVCGIQNSVSANRSDPVTLDVLYGPDTPIISPPDSSYLSGADLNLSCHSASNPSPQY
                     SWRINGIPQQHTQVLFIAKITPNNNGTYACFVSNLATGRNNSIVKSITVSASGTSPGL
                     SAGATVGIMIGVLVGVALI"
ORIGIN

Alignment Scores:
Pred. No.:                2.35          Length:          2106
Score:                    43.00         Matches:         8
Percent Similarity:       100.00%       Conservative:    1
Best Local Similarity:    88.89%        Mismatches:      0
Query Match:              95.56%        Indels:          0
DB:                       6             Gaps:            0

US-10-725-373-3 (1-9) x AX133657 (1-2106)

Qy          1 TyrLeuSerGlyAlaAspIleAsnLeu 9
              ||||||||||||||||||:::||||||
Db       1810 TACCTTTCGGGAGCGGACCTCAACCTC 1836

RESULT 6
AX192349
LOCUS       AX192349               2106 bp    DNA     linear   PAT 15-AUG-2001
DEFINITION  Sequence 3 from Patent WO0149317.
ACCESSION   AX192349
VERSION     AX192349.1  GI:15210326
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1
  AUTHORS   Emtage,P., Barber,B.H., Sambhara,S. and Sia,C.D.
  TITLE     Enhancing the immune response to an antigen by presensitizing with
            an inducing agent prior to immunizing with the inducing agent and
            the antigen
  JOURNAL   Patent: WO 0149317-A 3 12-JUL-2001;
            Aventis Pasteur Limited (CA)
FEATURES             Location/Qualifiers
     source          1. .2106
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
     CDS             1. .2106
                     /note="unnamed protein product; modified CEA"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAC51319.1"
                     /db_xref="GI:15210327"
                     /translation="MESPSAPPHRWCIPWQRLLLTASLLTFWNPPTTAKLTIESTPFN
                     VAEGKEVLLLVHNLPQHLFGYSWYKGERVDGNRQIIGYVIGTQQATPGPAYSGREIIY
                     PNASLLIQNIIQNDTGFYTLHVIKSDLVNEEATGQFRVYPELPKPSISSNNSKPVEDK
                     DAVAFTCEPETQDATYLWWVNNQSLPVSPRLQLSNGNRTLTLFNVTRNDTASYKCETQ
                     NPVSARRSDSVILNVLYGPDAPTISPLNTSYRSGENLNLSCHAASNPPAQYSWFVNGT
                     FQQSTQELFIPNITVNNSGSYTCQAHNSDTGLNRTTVTTITVYEPPKPFITSNNSNPV
                     EDEDAVALTCEPEIQNTTYLWWVNNQSLPVSPRLQLSNDNRTLTLLSVTRNDVGPYEC
                     GIQNELSVDHSDPVILNVLYGPDDPTISPSYTYYRPGVNLSLSCHAASNPPAQYSWLI
                     DGNIQQHTQELFISNITEKNSGLYTCQANNSASGHSRTTVKTITVSAELPKPSISSNN
                     SKPVEDKDAVAFTCEPEAQNTTYLWWVNGQSLPVSPRLQLSNGNRTLTLFNVTRNDAR
                     AYVCGIQNSVSANRSDPVTLDVLYGPDTPIISPPDSSYLSGADLNLSCHSASNPSPQY
                     SWRINGIPQQHTQVLFIAKITPNNNGTYACFVSNLATGRNNSIVKSITVSASGTSPGL
                     SAGATVGIMIGVLVGVALI"
ORIGIN

Alignment Scores:
Pred. No.:                2.35          Length:          2106
Score:                    43.00         Matches:         8
Percent Similarity:       100.00%       Conservative:    1
Best Local Similarity:    88.89%        Mismatches:      0
Query Match:              95.56%        Indels:          0
DB:                       6             Gaps:            0

US-10-725-373-3 (1-9) x AX192349 (1-2106)

Qy          1 TyrLeuSerGlyAlaAspIleAsnLeu 9
              ||||||||||||||||||:::||||||
Db       1810 TACCTTTCGGGAGCGGACCTCAACCTC 1836

RESULT 7
AX393888
LOCUS       AX393888               2106 bp    DNA     linear   PAT 23-MAR-2002
DEFINITION  Sequence 2 from Patent WO0210379.
ACCESSION   AX393888
VERSION     AX393888.1  GI:19701852
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1
  AUTHORS   Berinstein,N., Tartaglia,J., Tine,J.A., Panicali,D.L., Gritz,L. and
            Schlom,J.
  TITLE     Modified cea and uses thereof
  JOURNAL   Patent: WO 0210379-A 2 07-FEB-2002;
            Aventis Pasteur Limited (CA) ; Therion Biologics (US) ; National
            Cancer Institute (US)
FEATURES             Location/Qualifiers
     source          1. .2106
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="Cea modified polypeptide"
ORIGIN

Alignment Scores:
Pred. No.:                2.35          Length:          2106
Score:                    43.00         Matches:         8
Percent Similarity:       100.00%       Conservative:    1
Best Local Similarity:    88.89%        Mismatches:      0
Query Match:              95.56%        Indels:          0
DB:                       6             Gaps:            0

US-10-725-373-3 (1-9) x AX393888 (1-2106)

Qy          1 TyrLeuSerGlyAlaAspIleAsnLeu 9
              ||||||||||||||||||:::||||||
Db       1810 TACCTTTCGGGAGCGGACCTCAACCTC 1836
```

```
RESULT 8
AC136727
LOCUS       AC136727              151138 bp    DNA     linear   HTG 23-APR-2003
DEFINITION  Mus musculus clone RP23-72D18, WORKING DRAFT SEQUENCE, 7 unordered
            pieces.
ACCESSION   AC136727
VERSION     AC136727.3  GI:30017820
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 151138)
  AUTHORS   Birren,B., Nusbaum,C. and Lander,E.
  TITLE     Mus musculus, clone RP23-72D18
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 151138)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L., Boukhgalter,B.,
            Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
            Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hafez,N.,
            Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
            Lindblad-Toh,K., Liu,G., MacLean,C., Macdonald,P., Major,J.,
            Matthews,C., McCarthy,M., Meldrim,J., Meneus,L., Mihova,T.,
            Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
            Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J.,
            Peterson,K., Phunkhang,P., Pierre,N., Raymond,C., Retta,R.,
            Rise,C., Rogov,P., Roman,J., Roy,A., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Talamas,J., Tesfaye,S., Theodore,J., Topham,K.,
            Travers,M., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X.,
            Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-NOV-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 151138)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
            Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
            Boguslavkiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y.,
            Collymore,A., Cook,A., Cooke,P., Corum,B., DeArellano,K.,
            Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,
            Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,
            Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,
            Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
            Lindblad-Toh,K., Liu,G., Lui,A., Mabbitt,R., MacLean,C.,
            Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
            Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J.,
            Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
            Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
            Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C.,
            Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
            Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M.,
            Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
            Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-APR-2003) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Apr 17, 2003 this sequence version replaced gi:27311396.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L26304
                Center clone name: 72_D_18
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 7 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    27871: contig of 27871 bp in length
            *    27872    27971: gap of 100 bp
            *    27972    29011: contig of 1040 bp in length
            *    29012    29111: gap of 100 bp
            *    29112    39931: contig of 10820 bp in length
            *    39932    40031: gap of 100 bp
            *    40032    52297: contig of 12266 bp in length
            *    52298    52397: gap of 100 bp
            *    52398    72167: contig of 19770 bp in length
            *    72168    72267: gap of 100 bp
            *    72268    93947: contig of 21680 bp in length
            *    93948    94047: gap of 100 bp
            *    94048   151138: contig of 57091 bp in length.
FEATURES             Location/Qualifiers
     source          1. .151138
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /clone="RP23-72D18"
                     /clone_lib="RPCI-23 Female Mouse BAC"
     misc_feature    1. .27871
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    27972. .29011
                     /note="assembly_fragment"
     misc_feature    29112. .39931
                     /note="assembly_fragment"
     misc_feature    40032. .52297
                     /note="assembly_fragment"
     misc_feature    52398. .72167
                     /note="assembly_fragment"
     misc_feature    72268. .93947
                     /note="assembly_fragment"
     misc_feature    94048. .151138
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
ORIGIN

Alignment Scores:
Pred. No.:              798             Length:         151138
Score:                  42.00           Matches:        8
Percent Similarity:     100.00%         Conservative:   1
Best Local Similarity:  88.89%          Mismatches:     0
Query Match:            93.33%          Indels:         0
DB:                     2               Gaps:           0

US-10-725-373-3 (1-9) x AC136727 (1-151138)

Qy          1 TyrLeuSerGlyAlaAspIleAsnLeu 9
              ||||||||||||:::|||||||||||
Db      50277 TATCTCAGTGGGTCTGACATCAACCTT 50303

RESULT 9
AC149846/c
LOCUS       AC149846              185385 bp    DNA     linear   HTG 06-AUG-2004
DEFINITION  Papio anubis clone RP41-153B23, WORKING DRAFT SEQUENCE, 4 ordered
            pieces.
ACCESSION   AC149846
VERSION     AC149846.2  GI:51011162
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT.
SOURCE      Papio anubis (olive baboon)
```

```
  ORGANISM  Papio anubis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
            Cercopithecinae; Papio.
REFERENCE   1  (bases 1 to 185385)
  AUTHORS   Antonellis,A., Ayele,K., Benjamin,B., Blakesley,R.W.,
            Bouffard,G.G., Brinkley,C., Brooks,S., Chu,G., Coleman,B.,
            Coleman,H., Daki,N., Engle,J., Guan,X., Gupta,J., Haghighi,P.,
            Han,J., Hansen,N., Ho,S.-L., Hu,P., Hurle,B., Idol,J.R., Jones,C.,
            Karlins,E., Kim,H., Kwong,P., Laric,P., Larson,S., Lee-Lin,S.-Q.,
            Legaspi,R., Madden,M., Maduro,Q.L., Maduro,V.B., Margulies,E.H.,
            Masiello,C., Maskeri,B., McDowell,J., Mullikin,J.C., Paguirigan,C.,
            Park,M., Portnoy,M.E., Prasad,A., Puri,O., Reddix-Dugue,N.,
            Schandler,K., Schueler,M.G., Shah,K., Sison,C., Stantripop,S.,
            Thomas,J.W., Thomas,P.J., Tsipouri,V., Vogt,J.L., Wetherby,K.D.,
            Young,A. and Green,E.D.
  TITLE     NISC Comparative Sequencing Initiative
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 185385)
  AUTHORS   Green,E.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JUN-2004) NIH Intramural Sequencing Center, 8717
            Grovemont Circle, Gaithersburg, MD 20877, USA
REFERENCE   3  (bases 1 to 185385)
  AUTHORS   Green,E.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-AUG-2004) NIH Intramural Sequencing Center, 8717
            Grovemont Circle, Gaithersburg, MD 20877, USA
COMMENT     On Aug 6, 2004 this sequence version replaced gi:49065688.
            -------------- Genome Center
                 Center: NIH Intramural Sequencing Center
                 Center code: NISC
                 Web site: http://www.nisc.nih.gov
                 Contact: nisc_zoo@nhgri.nih.gov
            -------------- Project Information
                 Center project name: hpx
                 Center clone name: 153B23

            The sequence data in this record represents an 'enhanced'
            version of a Phase 2 submission. Specifically, the indicated
            order and orientation of each sequence contig has been
            established using one or more of the following: read-pair
            data from individual subclones, overlaps with neighboring
            clones, alignment with available reference sequence (e.g.,
            human), and/or confirmation by PCR testing. In addition,
            the sequence assembly is based on at least 8X average
            coverage in Q20 bases and has been reviewed to rule out
            gross misassemblies, the low-quality ends of sequence
            contigs have been trimmed away, and each base is associated
            with a Phrap-derived quality score.
            -------------- Summary Statistics
                 Sequencing vector: plasmid; n/a; 100% of reads
                 Chemistry: Dye-terminator Big Dye; 100% of reads
                 Assembly program: Phrap; version 0.990319
                 Consensus quality: 183668 bases at least Q40
                 Consensus quality: 184188 bases at least Q30
                 Consensus quality: 184625 bases at least Q20
                 Insert size: 180000; agarose-fp
                 Insert size: 185085; sum-of-contigs
                 Quality coverage: 10.31x in Q20 bases; agarose-fp
                 Quality coverage: 10.03x in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 4 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *       1   104105: contig of 104105 bp in length
            *  104106   104205: gap of unknown length
            *  104206   109148: contig of 4943 bp in length
            *  109149   109248: gap of unknown length
            *  109249   130407: contig of 21159 bp in length
            *  130408   130507: gap of unknown length
            *  130508   185385: contig of 54878 bp in length.
FEATURES             Location/Qualifiers
     source          1. .185385
                     /organism="Papio anubis"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9555"
                     /clone="RP41-153B23"
                     /clone_lib="RP41"
                     /note="BAC resource: http://bacpac.chori.org/"
     misc_feature    1. .104105
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    1. .57045
                     /note="clone overlaps with GenBank Accession Number
                     AC150306 clone RP41-49E22 (center project name hqg)"
     misc_feature    104206. .109148
                     /note="assembly_fragment"
     misc_feature    109249. .130407
                     /note="assembly_fragment"
     misc_feature    114267. .185385
                     /note="clone overlaps with GenBank Accession Number
                     AC149567 clone RP41-462A8 (center project name hpy)"
     misc_feature    130508. .185385
                     /note="assembly_fragment
                     missing T7 clone end on 3' end of insert"
ORIGIN

Alignment Scores:
Pred. No.:                1.02e+03        Length:          185385
Score:                    42.00           Matches:         8
Percent Similarity:       100.00%         Conservative:    1
Best Local Similarity:    88.89%          Mismatches:      0
Query Match:              93.33%          Indels:          0
DB:                       2               Gaps:            0

US-10-725-373-3 (1-9) x AC149846 (1-185385)

Qy          1 TyrLeuSerGlyAlaAspIleAsnLeu 9
              ||||||:::||||||||||||||||||
Db      74684 TATCTCAATGGGGCTGATATAAATTTG 74658

RESULT 10
AY115485
LOCUS       AY115485                8239 bp    DNA     linear   PLN 11-FEB-2004
DEFINITION  Zea mays anthocyanin biosynthetic gene regulator PAC1 (pac1) gene,
            complete cds.
ACCESSION   AY115485
VERSION     AY115485.1  GI:37544702
KEYWORDS    .
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 8239)
  AUTHORS   Carey,C.C., Strahle,J.T., Selinger,D.A. and Chandler,V.L.
  TITLE     Mutations in the pale aleurone color1 Regulatory Gene of the Zea
            mays Anthocyanin Pathway Have Distinct Phenotypes Relative to the
            Functionally Similar TRANSPARENT TESTA GLABRA1 Gene in Arabidopsis
            thaliana
  JOURNAL   Plant Cell 16 (2), 450-464 (2004)
   PUBMED   14742877
REFERENCE   2  (bases 1 to 8239)
  AUTHORS   Carey,C.C., Chandler,V.L. and Strahle,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-MAY-2002) Plant Sciences, University of Arizona, 303
            Forbes Building, Tucson, AZ 85721, USA
FEATURES             Location/Qualifiers
```

```
     source          1. .8239
                     /organism="Zea mays"
                     /mol_type="genomic DNA"
                     /cultivar="B73"
                     /db_xref="taxon:4577"
                     /chromosome="5"
                     /map="within 3.5 cm of phi087 SSR marker (phi087 is at
                     coordinate 103.3 on chromosome 5 on Pioneer Composite 1999
                     map)"
                     /clone="ZMMBBb0124P19"
     gene            <3320. .7854
                     /gene="pac1"
     mRNA            join(<3320. .4385,7403. .7854)
                     /gene="pac1"
                     /product="anthocyanin biosynthetic gene regulator PAC1"
     CDS             3320. .4381
                     /gene="pac1"
                     /note="WD40 repeat protein; required for transcriptional
                     activation of anthocyanin biosynthetic genes; possible
                     transcriptional regulator"
                     /codon_start=1
                     /product="anthocyanin biosynthetic gene regulator PAC1"
                     /protein_id="AAM76742.1"
                     /db_xref="GI:37544703"
                     /translation="MDPPKPPSSVASSSGPETPNPHAFTCELPHSIYALAFSPVAPVL
                     ASGSFLEDLHNRVSLLSFDPVRPSAASFRALPALSFDHPYPPTKLQFNPRAAAPSLLA
                     SSADTLRIWHTPLDDLSDTAPAPELRSVLDNRKASSEFCAPLTSFDWNEVEPRRIGTA
                     SIDTTCTVWDIDRGVVETQLIAHDKAVHDIAWGEAGVFASVSADGSVRVFDLRDKEHS
                     TIVYESPRPDTPLLRLAWNRSDLRYMAALLMDSSAVVVLDIRAPGVPVAELHRHRACA
                     NAVAWAPQATRHLCSAGDDGQALIWELPETAAAVPAEGIDPVLVYDAGAEINQLQWAA
                     AHPDWMAIAFENKVQLLRV"
     unsure          3772
                     /gene="pac1"
                     /note="sequencing ambiguity; may encode Glu or Asp"
     3'UTR           join(4382. .4385,7403. .7854)
                     /gene="pac1"
ORIGIN

Alignment Scores:
Pred. No.:                41.1          Length:          8239
Score:                    41.00         Matches:         7
Percent Similarity:       100.00%       Conservative:    2
Best Local Similarity:    77.78%        Mismatches:      0
Query Match:              91.11%        Indels:          0
DB:                       8             Gaps:            0

US-10-725-373-3 (1-9) x AY115485 (1-8239)

Qy          1 TyrLeuSerGlyAlaAspIleAsnLeu 9
              ||||||||||||:::|||:::||||||
Db       2313 TATTTAAGTGGCTCTGATGTTAATTTA 2339

RESULT 11
AL161654/c
LOCUS       AL161654               159231 bp    DNA     linear   PRI 18-JUL-2001
DEFINITION  Human DNA sequence from clone RP11-59G22 on chromosome 10, complete
            sequence.
ACCESSION   AL161654
VERSION     AL161654.14  GI:14970795
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 159231)
  AUTHORS   Sycamore,N.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Jul 19, 2001 this sequence version replaced gi:14625538.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. The following
            abbreviations are used to associate primary accession numbers given
            in the feature table with their source databases: Em:, EMBL; Sw:,
            SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
            database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 10, constructed by the Sanger Centre Chromosome 10
            Mapping Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr10
            RP11-59G22 is from the library RPCI-11.1 constructed by the group
            of Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6
            This sequence is the entire insert of clone RP11-59G22 The true
            right end of clone RP13-236A4 is at 127080 in this sequence. The
            true right end of clone RP11-561H23 is at 150489 in this sequence.
FEATURES             Location/Qualifiers
     source          1. .159231
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="10"
                     /clone="RP11-59G22"
                     /clone_lib="RPCI-11.1"
     repeat_region   522. .785
                     /note="L1M3b repeat: matches -26. .234 of consensus"
     repeat_region   942. .1304
                     /note="THE1B repeat: matches 1. .362 of consensus"
     repeat_region   1307. .2879
                     /note="THE1B-INTERNAL repeat: matches 1. .1580 of
                     consensus"
     repeat_region   2880. .3244
                     /note="THE1B repeat: matches 1. .364 of consensus"
     repeat_region   3245. .3711
                     /note="L1M4 repeat: matches 2407. .2876 of consensus"
     repeat_region   3783. .5500
                     /note="L1PA7 repeat: matches 4425. .6140 of consensus"
     repeat_region   5501. .7338
                     /note="L1MB2 repeat: matches 4261. .6097 of consensus"
     repeat_region   7902. .7949
                     /note="MIR repeat: matches 65. .109 of consensus"
     repeat_region   7954. .8114
                     /note="LTR45 repeat: matches 349. .512 of consensus"
     repeat_region   8416. .8571
                     /note="MIR repeat: matches 97. .251 of consensus"
     repeat_region   8513. .8749
                     /note="MIR repeat: matches 14. .256 of consensus"
     repeat_region   10278. .10896
                     /note="L1ME3 repeat: matches 5498. .6134 of consensus"
     repeat_region   10940. .11307
                     /note="L1MA8 repeat: matches 5931. .6286 of consensus"
     repeat_region   11376. .12159
                     /note="L1ME3 repeat: matches 4468. .5347 of consensus"
     repeat_region   12166. .12382
                     /note="L1M4 repeat: matches 2298. .2523 of consensus"
     repeat_region   12538. .12791
                     /note="LTR16C repeat: matches 136. .387 of consensus"
     repeat_region   12802. .13208
                     /note="L1P4 repeat: matches 5189. .5607 of consensus"
     repeat_region   13207. .15888
                     /note="L1P repeat: matches 1941. .4629 of consensus"
     repeat_region   15889. .17370
```

```
                     /note="L1PA3 repeat: matches 4664. .6144 of consensus"
     repeat_region   17411. .17905
                     /note="L1PA13 repeat: matches 5593. .6099 of consensus"
     repeat_region   17982. .18683
                     /note="L1M4 repeat: matches 2380. .3111 of consensus"
     repeat_region   18762. .20758
                     /note="L1MD2 repeat: matches 4313. .6325 of consensus"
     repeat_region   21404. .21449
                     /note="23 copies 2 mer at 76% conserved"
     repeat_region   21884. .22601
                     /note="L1MB7 repeat: matches 5430. .6171 of consensus"
     repeat_region   22673. .22830
                     /note="L1ME3 repeat: matches 5910. .6072 of consensus"
     repeat_region   22967. .23112
                     /note="73 copies 2 mer tt 59% conserved"
     repeat_region   23294. .23600
                     /note="AluJo repeat: matches 1. .308 of consensus"
     repeat_region   24141. .24220
                     /note="40 copies 2 mer tt 63% conserved"
     repeat_region   24597. .24681
                     /note="L2 repeat: matches 2664. .2749 of consensus"
     repeat_region   24860. .25065
                     /note="L1ME3 repeat: matches 5734. .5949 of consensus"
     repeat_region   25236. .25417
                     /note="L2 repeat: matches 1949. .2140 of consensus"
     repeat_region   25500. .25894
                     /note="L2 repeat: matches 1215. .1633 of consensus"
     repeat_region   27330. .27644
                     /note="AluSg repeat: matches 1. .310 of consensus"
     repeat_region   28380. .28575
                     /note="MIR repeat: matches 31. .259 of consensus"
     repeat_region   29470. .32026
                     /note="L1MB2 repeat: matches 3607. .6162 of consensus"
     repeat_region   32027. .32321
                     /note="AluSg repeat: matches 3. .297 of consensus"
     repeat_region   32322. .32541
                     /note="L1MB2 repeat: matches 3391. .3607 of consensus"
     repeat_region   33157. .34356
                     /note="MER52A repeat: matches 50. .1238 of consensus"
     repeat_region   34357. .34711
                     /note="THE1B repeat: matches 1. .364 of consensus"
     repeat_region   34712. .36326
                     /note="THE1B-INTERNAL repeat: matches 1. .1580 of
                     consensus"
     repeat_region   36327. .36689
                     /note="THE1B repeat: matches 1. .364 of consensus"
     repeat_region   36690. .37023
                     /note="MER52A repeat: matches 1238. .1755 of consensus"
     repeat_region   38016. .38117
                     /note="L2 repeat: matches 1494...1599 of consensus"
     repeat_region   38118. .38190
                     /note="7SK repeat: matches 121. .193 of consensus"
     repeat_region   38227. .38352
                     /note="7SK repeat: matches 2. .125 of consensus"
     repeat_region   38355. .38745
                     /note="L2 repeat: matches 1585. .1981 of consensus"
     repeat_region   40879. .41025
                     /note="MLT1E repeat: matches 413. .559 of consensus"
     repeat_region   41027. .41142
                     /note="FLAM_C repeat: matches 1. .116 of consensus"
     repeat_region   41322. .41551
                     /note="MLT1E repeat: matches 180. .419 of consensus"
     repeat_region   42225. .42306
                     /note="MLT1E repeat: matches 1. .87 of consensus"
     repeat_region   42426. .42795
                     /note="MLT1E repeat: matches 180. .560 of consensus"
     repeat_region   43403. .43493
                     /note="MER1B repeat: matches 1. .95 of consensus"
     repeat_region   43494. .43848
                     /note="THE1C repeat: matches 1. .371 of consensus"
     repeat_region   43849. .44110
                     /note="MER1B repeat: matches 95. .337 of consensus"
     repeat_region   44493. .44547
                     /note="L2 repeat: matches 2653. .2710 of consensus"
     repeat_region   46458. .47256
                     /note="L1MC4 repeat: matches 7167. .7968 of consensus"
     repeat_region   50004. .50037
                     /note="17 copies 2 mer tt 94% conserved"
     repeat_region   50668. .50940
                     /note="LTR16C repeat: matches 1. .282 of consensus"
     repeat_region   50973. .51208
                     /note="LTR24 repeat: matches 250. .489 of consensus"
     repeat_region   51396. .51595
                     /note="LTR24 repeat: matches 1. .198 of consensus"
     repeat_region   51882. .52432
                     /note="L2 repeat: matches 1597. .2185 of consensus"
     repeat_region   54848. .55146
                     /note="AluSg repeat: matches 3. .301 of consensus"
     repeat_region   56164. .56473
                     /note="AluY repeat: matches 1. .311 of consensus"
     repeat_region   57551. .57853
                     /note="AluJo repeat: matches 3. .290 of consensus"
     repeat_region   59379. .59507
                     /note="MIR repeat: matches 14. .143 of consensus"
     repeat_region   59514. .60217
                     /note="L1MB7 repeat: matches 4968. .5704 of consensus"
     repeat_region   60219. .60535
                     /note="L1MB7 repeat: matches 5845. .6173 of consensus"
     repeat_region   63607. .64353
                     /note="L1MB5 repeat: matches 5319. .6063 of consensus"
     repeat_region   64363. .64491
                     /note="L1MC1 repeat: matches 6010. .6139 of consensus"
     repeat_region   64490. .64675
                     /note="L1M4 repeat: matches 5145. .5332 of consensus"
     repeat_region   64676. .64962
                     /note="AluJo repeat: matches 1. .284 of consensus"
     repeat_region   64963. .65710
                     /note="L1M4 repeat: matches 4409. .5145 of consensus"

Alignment Scores:
Pred. No.:                1.54e+03        Length:          159231
Score:                    41.00           Matches:         8
Percent Similarity:       100.00%         Conservative:    1
Best Local Similarity:    88.89%          Mismatches:      0
Query Match:              91.11%          Indels:          0
DB:                       9               Gaps:            0

US-10-725-373-3 (1-9) x AL161654 (1-159231)

Qy          1 TyrLeuSerGlyAlaAspIleAsnLeu 9
              ||||||||||||||:::|||||||||
Db     137882 TATCTGAGTGGTGCAGAGATAAACTTG 137856

RESULT 12
AL591477/c
LOCUS       AL591477              168608 bp    DNA     linear   HTG 19-DEC-2001
DEFINITION  Homo sapiens chromosome 10 clone RP13-112H19, WORKING DRAFT
            SEQUENCE, 3 unordered pieces.
ACCESSION   AL591477
VERSION     AL591477.2  GI:17973980
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Burton,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-DEC-2001) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Dec 20, 2001 this sequence version replaced gi:14141526.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
```

```
          Contact: humquery@sanger.ac.uk
          -------------- Project Information
          Center project name: bB112H19
          -------------- Summary Statistics
          Assembly program: XGAP4; version 4.5
          Sequencing vector: plasmid; L08752; 100% of reads
          Chemistry: Dye-terminator Big Dye; 100% of reads
          Consensus quality: 167544 bases at least Q40
          Consensus quality: 167716 bases at least Q30
          Consensus quality: 167994 bases at least Q20
          Insert size: 168408; sum-of-contigs
          Insert size: 152446; agarose-fp
          Quality coverage: 9.85x in Q20 bases; sum-of-contigs Quality
          coverage: 10.88x in Q20 bases; agarose-fp
          --------------.
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 3 contigs. The true order of the pieces
          * is not known and their order in this sequence record is
          * arbitrary. Gaps between the contigs are represented as
          * runs of N, but the exact sizes of the gaps are unknown.
          * This record will be updated with the finished sequence
          * as soon as it is available and the accession number will
          * be preserved.
          *        1    35137: contig of 35137 bp in length
          *    35138    35237: gap of 100 bp
          *    35238    60541: contig of 25304 bp in length
          *    60542    60641: gap of 100 bp
          *    60642   168608: contig of 107967 bp in length.
FEATURES             Location/Qualifiers
     source          1. .168608
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="10"
                     /clone="RP13-112H19"
                     /clone_lib="RPCI-13.1"
     misc_feature    1. .35137
                     /note="assembly_fragment:00872
                     clone_end:T7
                     vector_side:left"
     misc_feature    35238. .60541
                     /note="assembly_fragment:01936"
     misc_feature    60642. .168608
                     /note="assembly_fragment:00906
                     clone_end:SP6
                     vector_side:right"
ORIGIN

Alignment Scores:
Pred. No.:              1.66e+03        Length:         168608
Score:                  41.00           Matches:        8
Percent Similarity:     100.00%         Conservative:   1
Best Local Similarity:  88.89%          Mismatches:     0
Query Match:            91.11%          Indels:         0
DB:                     2               Gaps:           0

US-10-725-373-3 (1-9) x AL591477 (1-168608)

Qy          1 TyrLeuSerGlyAlaAspIleAsnLeu 9
              ||||||||||||||:::|||||||||
Db      36233 TATCTGAGTGGTGCAGAGATAAACTTG 36207

RESULT 13
AR560607
LOCUS       AR560607                  27 bp    DNA     linear   PAT 08-OCT-2004
DEFINITION  Sequence 11 from patent US 6756038.
ACCESSION   AR560607
VERSION     AR560607.1  GI:53972928
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 27)
  AUTHORS   Schlom,J., Barzaga,E. and Zaremba,S.
  TITLE     Agonist and antagonist peptides of carcinoembryonic antigen (CEA)
  JOURNAL   Patent: US 6756038-A 11 29-JUN-2004;
FEATURES             Location/Qualifiers
     source          1. .27
                     /organism="unknown"
                     /mol_type="genomic DNA"
ORIGIN

Alignment Scores:
Pred. No.:              0.068           Length:         27
Score:                  40.00           Matches:        8
Percent Similarity:     100.00%         Conservative:   1
Best Local Similarity:  88.89%          Mismatches:     0
Query Match:            88.89%          Indels:         0
DB:                     6               Gaps:           0

US-10-725-373-3 (1-9) x AR560607 (1-27)

Qy          1 TyrLeuSerGlyAlaAspIleAsnLeu 9
              ||||||||||||||:::|||||||||
Db          1 TACCTTTCGGGAGCGAACATCAACCTC 27

RESULT 14
BD131678
LOCUS       BD131678                  27 bp    DNA     linear   PAT 18-SEP-2002
DEFINITION  Carcinoembryonic antigen (CEA) agonist and antagonist peptides.
ACCESSION   BD131678
VERSION     BD131678.1  GI:23226623
KEYWORDS    JP 2002500002-A/4.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 27)
  AUTHORS   Schlom,J., Barzaga,E. and Zaremba,S.
  TITLE     Carcinoembryonic antigen (CEA) agonist and antagonist peptides
  JOURNAL   Patent: JP 2002500002-A 4 08-JAN-2002;
            THE UNITED STATES OF AMERICA
COMMENT     OS   Homo sapiens (human)
            PN   JP 2002500002-A/4
            PD   08-JAN-2002
            PF   22-SEP-1998 JP 2000516030
            PR   10-OCT-1997 US    60/061589
            PI   JEFFREY SCHLOM,ELENE BARZAGA,SAM ZAREMBA
            PC   C12N15/09,A61K38/00,A61K45/00,A61K48/00,A61P35/00,A61P37/02,
            PC   A61P43/00,
            PC   C07K14/705,C07K16/28,C12N5/10,C12N15/00,A61K37/02,C12N5/00 CC
             Carcinoembryonic antigen (CEA) agonist and antagonist peptides FH
             Key             Location/Qualifiers
            FT   source          1. .27
            FT                   /organism='Homo sapiens (human)'.
FEATURES             Location/Qualifiers
     source          1. .27
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
ORIGIN

Alignment Scores:
Pred. No.:              0.068           Length:         27
Score:                  40.00           Matches:        8
Percent Similarity:     100.00%         Conservative:   1
Best Local Similarity:  88.89%          Mismatches:     0
Query Match:            88.89%          Indels:         0
DB:                     6               Gaps:           0

US-10-725-373-3 (1-9) x BD131678 (1-27)

Qy          1 TyrLeuSerGlyAlaAspIleAsnLeu 9
              ||||||||||||||:::|||||||||
Db          1 TACCTTTCGGGAGCGAACATCAACCTC 27
```

```
RESULT 15
CQ869594/c
LOCUS       CQ869594              108845 bp    DNA     linear   PAT 13-SEP-2004
DEFINITION  Sequence 15 from Patent WO2004074320.
ACCESSION   CQ869594
VERSION     CQ869594.1  GI:51999464
KEYWORDS    .
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Morris,D.W., Morris,D.W. and Malandro,M.S.
  TITLE     Novel therapeutic targets in cancer
  JOURNAL   Patent: WO 2004074320-A 15 02-SEP-2004;
            Sagres Discovery, Inc. (US)
FEATURES             Location/Qualifiers
     source          1. .108845
                     /organism="Mus musculus"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:10090"
ORIGIN

Alignment Scores:
Pred. No.:              1.76e+03        Length:         108845
Score:                  40.00           Matches:        7
Percent Similarity:     100.00%         Conservative:   2
Best Local Similarity:  77.78%          Mismatches:     0
Query Match:            88.89%          Indels:         0
DB:                     6               Gaps:           0

US-10-725-373-3 (1-9) x CQ869594 (1-108845)

Qy          1 TyrLeuSerGlyAlaAspIleAsnLeu 9
              ||||||||||:::|||||||||:::
Db     105618 TATTTGAGTGGGTCAGATATTAACATA 105592


Search completed: May 17, 2005, 19:12:53
Job time : 1340.5 secs
```

OM protein - nucleic search, using frame_plus_p2n model

Run on: May 17, 2005, 16:32:09 ; Search time 1297.5 Seconds
(without alignments)
336.106 Million cell updates/sec

Title: US-10-725-373-5
Perfect score: 48
Sequence: 1 YLSGACLNL 9

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/US10725373/runat_17052005_071020_16185/app_query.fasta_1.796
-DB=GenEmbl -QFMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10725373_@CGN_1_1_6148_@runat_17052005_071020_16185 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	48	100.0	27	6	AR560608	AR560608 Sequence
	2	48	100.0	27	6	BD131679	BD131679 Carcinoem
	3	46	95.8	173756	2	AC105170	AC105170 Mus muscu
	4	46	95.8	252143	2	AC095437	AC095437 Rattus no
c	5	46	95.8	297470	2	AC125817	AC125817 Rattus no
c	6	44	91.7	114518	2	CR753825	CR753825 Homo sapi
	7	44	91.7	163584	9	AC008056	AC008056 Homo sapi
	8	44	91.7	250810	2	AC103486	AC103486 Rattus no
c	9	44	91.7	268281	2	AC124920	AC124920 Rattus no
c	10	44	91.7	274993	2	AC133403	AC133403 Rattus no
c	11	43	89.6	161363	10	AL645535	AL645535 Mouse DNA
	12	43	89.6	164991	9	AC092023	AC092023 Homo sapi
c	13	43	89.6	171347	9	AC099776	AC099776 Homo sapi
c	14	42	87.5	95663	9	AC010247	AC010247 Homo sapi
	15	42	87.5	112427	8	AC122164	AC122164 Medicago
	16	42	87.5	114109	9	AP002456	AP002456 Homo sapi
	17	42	87.5	130442	6	CQ870468	CQ870468 Sequence
	18	42	87.5	161547	9	AP001929	AP001929 Homo sapi
	19	42	87.5	167108	2	AC068283	AC068283 Homo sapi
	20	42	87.5	189956	2	AC117866	AC117866 Rattus no
c	21	42	87.5	203068	10	AL807755	AL807755 Mouse DNA
c	22	42	87.5	213572	2	AC105567	AC105567 Rattus no
c	23	42	87.5	236314	2	AC117628	AC117628 Mus muscu
	24	42	87.5	246624	2	AC125751	AC125751 Rattus no
	25	42	87.5	349980	6	CQ869981	CQ869981 Sequence
	26	41	85.4	647	9	AF020692	AF020692 Saguinus
	27	41	85.4	1884	6	AX513524	AX513524 Sequence
	28	41	85.4	1899	6	AR359364	AR359364 Sequence
	29	41	85.4	1899	6	AX098631	AX098631 Sequence
	30	41	85.4	1899	6	AX513520	AX513520 Sequence
	31	41	85.4	1899	6	AX513522	AX513522 Sequence
	32	41	85.4	1902	1	AY297091	AY297091 Bacillus
	33	41	85.4	1902	6	AR107058	AR107058 Sequence
	34	41	85.4	1902	6	AR107059	AR107059 Sequence
	35	41	85.4	1902	6	BD082244	BD082244 Bacillus
	36	41	85.4	1902	6	BD082245	BD082245 Bacillus
	37	41	85.4	1912	1	BACCRYIB2	M23724 B.thuringie
	38	41	85.4	1912	6	AR260589	AR260589 Sequence
	39	41	85.4	2281	1	AF164666	AF164666 Bacillus
c	40	41	85.4	2610	3	AK173522	AK173522 Ciona int
	41	41	85.4	2765	5	AF403114	AF403114 Gallus ga
	42	41	85.4	2812	5	AF403116	AF403116 Gallus ga
	43	41	85.4	2830	5	AF403118	AF403118 Gallus ga
	44	41	85.4	2877	5	AF403120	AF403120 Gallus ga
	45	41	85.4	2924	1	BTKCRY2B	X55416 B. thuringi

ALIGNMENTS

RESULT 1
AR560608
LOCUS AR560608 27 bp DNA linear PAT 08-OCT-2004
DEFINITION Sequence 12 from patent US 6756038.
ACCESSION AR560608
VERSION AR560608.1 GI:53972929
KEYWORDS .
SOURCE Unknown.
ORGANISM Unknown.
Unclassified.
REFERENCE 1 (bases 1 to 27)
AUTHORS Schlom,J., Barzaga,E. and Zaremba,S.
TITLE Agonist and antagonist peptides of carcinoembryonic antigen (CEA)
JOURNAL Patent: US 6756038-A 12 29-JUN-2004;
FEATURES Location/Qualifiers
source 1. .27
/organism="unknown"
/mol_type="genomic DNA"
ORIGIN

Alignment Scores:

Pred. No.:	0.237	Length:	27
Score:	48.00	Matches:	9
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	6	Gaps:	0

```
US-10-725-373-5 (1-9) x AR560608 (1-27)

Qy          1 TyrLeuSerGlyAlaCysLeuAsnLeu 9
              |||||||||||||||||||||||||||
Db          1 TACCTTTCGGGAGCGTGTCTCAACCTC 27

RESULT 2
BD131679
LOCUS       BD131679                  27 bp    DNA     linear   PAT 18-SEP-2002
DEFINITION  Carcinoembryonic antigen (CEA) agonist and antagonist peptides.
ACCESSION   BD131679
VERSION     BD131679.1  GI:23226624
KEYWORDS    JP 2002500002-A/5.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 27)
  AUTHORS   Schlom,J., Barzaga,E. and Zaremba,S.
  TITLE     Carcinoembryonic antigen (CEA) agonist and antagonist peptides
  JOURNAL   Patent: JP 2002500002-A 5 08-JAN-2002;
            THE UNITED STATES OF AMERICA
COMMENT     OS   Homo sapiens (human)
            PN   JP 2002500002-A/5
            PD   08-JAN-2002
            PF   22-SEP-1998 JP 2000516030
            PR   10-OCT-1997 US    60/061589
            PI   JEFFREY SCHLOM,ELENE BARZAGA,SAM ZAREMBA
            PC   C12N15/09,A61K38/00,A61K45/00,A61K48/00,A61P35/00,A61P37/02,
            PC   A61P43/00,
            PC   C07K14/705,C07K16/28,C12N5/10,C12N15/00,A61K37/02,C12N5/00 CC
             Carcinoembryonic antigen (CEA) agonist and antagonist peptides FH
             Key             Location/Qualifiers
            FT   source          1. .27
            FT                   /organism='Homo sapiens (human)'.
FEATURES             Location/Qualifiers
     source          1. .27
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
ORIGIN

Alignment Scores:
Pred. No.:              0.237          Length:         27
Score:                  48.00          Matches:        9
Percent Similarity:     100.00%        Conservative:   0
Best Local Similarity:  100.00%        Mismatches:     0
Query Match:            100.00%        Indels:         0
DB:                     6              Gaps:           0

US-10-725-373-5 (1-9) x BD131679 (1-27)

Qy          1 TyrLeuSerGlyAlaCysLeuAsnLeu 9
              |||||||||||||||||||||||||||
Db          1 TACCTTTCGGGAGCGTGTCTCAACCTC 27

RESULT 3
AC105170
LOCUS       AC105170              173756 bp    DNA     linear   HTG 28-FEB-2004
DEFINITION  Mus musculus chromosome 14 clone RP24-68I9 map 14, *** SEQUENCING
            IN PROGRESS ***, 2 ordered pieces.
ACCESSION   AC105170
VERSION     AC105170.5  GI:44681600
KEYWORDS    HTG; HTGS_PHASE2; HTGS_FULLTOP; HTGS_ACTIVEFIN.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 173756)
  AUTHORS   Birren,B., Nusbaum,C. and Lander,E.
  TITLE     Mus musculus chromosome 14, clone RP24-68I9
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 173756)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
            Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
            Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
            Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
            Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J.,
            Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-DEC-2001) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 173756)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
            Anderson,M., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
            Boguslavkiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y.,
            Collymore,A., Cook,A., Cooke,P., Corum,B., DeArellano,K.,
            Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,
            Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,
            Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,
            Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
            Lindblad-Toh,K., Liu,X., Lui,A., Mabbitt,R., MacLean,C.,
            Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
            Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J.,
            Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
            Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
            Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C.,
            Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
            Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M.,
            Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
            Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-FEB-2004) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Feb 28, 2004 this sequence version replaced gi:29029380.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L19291
                Center clone name: 68_I_9
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 2 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *        1    93193: contig of 93193 bp in length
            *    93194    93293: gap of 100 bp
```

```
          *    93294   173756: contig of 80463 bp in length.
FEATURES             Location/Qualifiers
     source          1. .173756
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /chromosome="14"
                     /map="14"
                     /clone="RP24-68I9"
                     /clone_lib="RPCI-24 Male Mouse BAC"
ORIGIN

Alignment Scores:
Pred. No.:                1.57e+03       Length:         173756
Score:                    46.00          Matches:        8
Percent Similarity:       100.00%        Conservative:   1
Best Local Similarity:    88.89%         Mismatches:     0
Query Match:              95.83%         Indels:         0
DB:                       2              Gaps:           0

US-10-725-373-5 (1-9) x AC105170 (1-173756)

Qy          1 TyrLeuSerGlyAlaCysLeuAsnLeu 9
              ||||||||||||||||||:::||||||
Db       9520 TACCTAAGTGGAGCATGCATTAACCTT 9546

RESULT 4
AC095437
LOCUS       AC095437               252143 bp    DNA     linear   HTG 09-MAY-2003
DEFINITION  Rattus norvegicus clone CH230-4H9, *** SEQUENCING IN PROGRESS ***,
            2 unordered pieces.
ACCESSION   AC095437
VERSION     AC095437.6  GI:30467826
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 252143)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 252143)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-SEP-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 252143)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-MAY-2003) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On May 9, 2003 this sequence version replaced gi:24941147.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GCHA
                Center clone name: CH230-4H9
            -------------- Summary Statistics
                Assembly program: Atlas;
                Consensus quality: 229361 bases at least Q40
                Consensus quality: 231822 bases at least Q30
                Consensus quality: 233149 bases at least Q20
                Estimated insert size: 249796; sum-of-contigs estimation
                Quality coverage: 6x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 2 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1   246919: contig of 246919 bp in length
            *   246920   247019: gap of unknown length
            *   247020   252143: contig of 5124 bp in length.
FEATURES             Location/Qualifiers
```

```
     source          1. .252143
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-4H9"
     misc_feature    complement(2228. .2885)
                     /note="clone_boundary
                     clone_end:Sp6
                     site:EcoRI
                     end_sequence:BH306797"
     misc_feature    247020. .248193
                     /note="wgs_contig"
ORIGIN

Alignment Scores:
Pred. No.:              2.19e+03        Length:         252143
Score:                  46.00           Matches:        8
Percent Similarity:     100.00%         Conservative:   1
Best Local Similarity:  88.89%          Mismatches:     0
Query Match:            95.83%          Indels:         0
DB:                     2               Gaps:           0

US-10-725-373-5 (1-9) x AC095437 (1-252143)

Qy          1 TyrLeuSerGlyAlaCysLeuAsnLeu 9
              ||||||||||||||||||:::|||||||
Db     132289 TACCTAAGCGGAGCATGCATTAACCTT 132315

RESULT 5
AC125817/c
LOCUS       AC125817               297470 bp    DNA     linear   HTG 13-NOV-2002
DEFINITION  Rattus norvegicus clone CH230-2N21, WORKING DRAFT SEQUENCE, 6
            unordered pieces.
ACCESSION   AC125817
VERSION     AC125817.3  GI:24635796
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 297470)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 297470)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUL-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 297470)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-NOV-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Nov 6, 2002 this sequence version replaced gi:22773028.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GDCE
                Center clone name: CH230-2N21
            -------------- Summary Statistics
                Assembly program: Phrap; version 0.990329
                Consensus quality: 218036 bases at least Q40
                Consensus quality: 220310 bases at least Q30
                Consensus quality: 221663 bases at least Q20
                Estimated insert size: 223274; sum-of-contigs estimation
                Quality coverage: 7x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 6 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1   282417: contig of 282417 bp in length
            *   282418   282517: gap of unknown length
```

```
            *   282518   284692: contig of 2175 bp in length
            *   284693   284792: gap of unknown length
            *   284793   285902: contig of 1110 bp in length
            *   285903   286002: gap of unknown length
            *   286003   287368: contig of 1366 bp in length
            *   287369   287468: gap of unknown length
            *   287469   289494: contig of 2026 bp in length
            *   289495   289594: gap of unknown length
            *   289595   297470: contig of 7876 bp in length.
FEATURES             Location/Qualifiers
     source          1. .297470
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-2N21"
ORIGIN

Alignment Scores:
Pred. No.:              2.54e+03        Length:         297470
Score:                  46.00           Matches:        8
Percent Similarity:     100.00%         Conservative:   1
Best Local Similarity:  88.89%          Mismatches:     0
Query Match:            95.83%          Indels:         0
DB:                     2               Gaps:           0

US-10-725-373-5 (1-9) x AC125817 (1-297470)

Qy          1 TyrLeuSerGlyAlaCysLeuAsnLeu 9
              ||||||||||||||||||:::||||||
Db     256539 TACCTAAGCGGAGCATGCATTAACCTT 256513

RESULT 6
CR753825/c
LOCUS       CR753825              114518 bp    DNA     linear   HTG 27-AUG-2004
DEFINITION  Homo sapiens chromosome 6 clone DASS-113F17, *** SEQUENCING IN
            PROGRESS ***, 6 unordered pieces.
ACCESSION   CR753825
VERSION     CR753825.2  GI:51591719
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 114518)
  AUTHORS   Sims,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-AUG-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Aug 27, 2004 this sequence version replaced gi:51571641.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            -------------- Project Information
            Center project name: bSS113F17
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 113479 bases at least Q40
            Consensus quality: 113691 bases at least Q30
            Consensus quality: 113811 bases at least Q20
            Insert size: 114018; sum-of-contigs
            Insert size: 127555; 13.8% error; agarose-fp
            Quality coverage: 7.24x in Q20 bases; sum-of-contigs Quality
            coverage: 6.58x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 6 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1    17537: contig of 17537 bp in length
            *   17538    17637: gap of 100 bp
            *   17638    27339: contig of 9702 bp in length
            *   27340    27439: gap of 100 bp
            *   27440    29573: contig of 2134 bp in length
            *   29574    29673: gap of 100 bp
            *   29674    33542: contig of 3869 bp in length
            *   33543    33642: gap of 100 bp
            *   33643    67478: contig of 33836 bp in length
            *   67479    67578: gap of 100 bp
            *   67579   114518: contig of 46940 bp in length.
FEATURES             Location/Qualifiers
     source          1. .114518
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /clone="DASS-113F17"
                     /clone_lib="DNA-arts-BAC.1-SSTO.1"
     misc_feature    1. .17537
                     /note="assembly_fragment:00126
                     fragment_chain:1"
     misc_feature    17638. .27339
                     /note="assembly_fragment:00045
                     fragment_chain:1"
     misc_feature    27440. .29573
                     /note="assembly_fragment:00005
                     fragment_chain:1"
     misc_feature    29674. .33542
                     /note="assembly_fragment:00021"
     misc_feature    33643. .67478
                     /note="assembly_fragment:00312.0"
     misc_feature    67579. .114518
                     /note="assembly_fragment:00667.0"
ORIGIN

Alignment Scores:
Pred. No.:              2.77e+03        Length:         114518
Score:                  44.00           Matches:        8
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            91.67%          Indels:         0
DB:                     2               Gaps:           0

US-10-725-373-5 (1-9) x CR753825 (1-114518)

Qy          1 TyrLeuSerGlyAlaCysLeuAsn 8
              ||||||||||||||||||||||||
Db      38582 TACCTTTCTGGTGCATGTTTGAAT 38559

RESULT 7
AC008056
LOCUS       AC008056              163584 bp    DNA     linear   PRI 20-NOV-1999
DEFINITION  Homo sapiens chromosome 14 clone RP11-242P2 containing gene for
            neurexin III gene, partial cds, complete sequence.
ACCESSION   AC008056
VERSION     AC008056.6  GI:6456144
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 163584)
  AUTHORS   Rowen,L., Madan,A., Qin,S., Abbasi,N., Baradarani,L., Birditt,B.,
            Bloom,S., Dors,M., Dickhoff,R., Fleetwood,P., Harrison,G.,
            James,R., Kaur,A., Madan,A., Owen,M.P., Ratcliffe,A., Shaffer,T.
            and Hood,L.
  TITLE     Sequencing of human chromosome 14 neurexin III gene
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 163584)
```

```
  AUTHORS   Rowen,L., Madan,A., Qin,S., Abbasi,N., Dors,M., Dickhoff,R.,
            Harrison,G., James,R., Loretz,C., Lasky,S., Madan,A., Prescott,S.,
            Ratcliffe,A., Shaffer,T. and Hood,L.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-JUL-1999) Multimegabase Sequencing Center, University
            of Washington, PO BOX 357730, Seattle, WA 98195, USA
REFERENCE   3  (bases 1 to 163584)
  AUTHORS   Rowen,L., Madan,A., Qin,S., Abbasi,N., Baradarani,L., Birditt,B.,
            Bloom,S., Dors,M., Dickhoff,R., Fleetwood,P., Harrison,G.,
            James,R., Kaur,A., Madan,A., Owen,M.P., Ratcliffe,A., Shaffer,T.
            and Hood,L.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-NOV-1999) Multimegabase Sequencing Center, University
            of Washington, PO BOX 357730, Seattle, WA 98195, USA
COMMENT     On Nov 20, 1999 this sequence version replaced gi:6274557.
            ------------ Genome Center
                Center: Multimegabase Sequencing Center
                Center code: UWMSC
                Web site: http://chroma.mbt.washington.edu/msg_www
                Contact: leerowen@u.washington.edu
            ------------ Summary Statistics
                Sequencing vector: pUC18; L08752
                Chemistry: Big Dye terminators and primers
                Assembly program: Phrap; version 0.990399
            ------------.
FEATURES             Location/Qualifiers
     source          1. .163584
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="14"
                     /map="14q31"
                     /clone="RP11-242P2"
                     /clone_lib="RPCI human BAC library 11"
                     /note="This clone overlaps RP11-45G3, Accession AC012099
                     and CTD-2294M16, AC007058"
     unsure          1450. .1620
                     /note="low quality data in microsatellite region"
     unsure          29640. .29680
                     /note="low quality data"
     unsure          62720. .62830
                     /note="low quality data"
     CDS             join(<100016. .100187,133868. .>134175)
                     /note="Intron-exon boundaries defined in relation to
                     KIAA0743"
                     /codon_start=1
                     /product="neurexin III"
                     /protein_id="AAF09143.1"
                     /db_xref="GI:6456145"
                     /translation="EQGKIGVVFNIGTVDISIKEERTPVNDGKYHVVRFTRNGGNATL
                     QVDNWPVNEHYPTGRQLTIFNTQAQIAIGGKDKGRLFQGQLSGLYYDGLKVLNMAAEN
                     NPNIKINGSVRLVGEVPSILGTTQTTSMPPEMSTTVMETTTTMATTTTRKNRSTASIQ
                     "
     unsure          110940. .111010
                     /note="low quality data"
     unsure          119050. .119080
                     /note="low quality data"
     misc_feature    146567. .163584
                     /note="overlap with CTD-2294M16, Accession AC007058"
     variation       146826
                     /note="242P2: g; 2294M16: a"
                     /replace="a"
     variation       147921
                     /note="242P2: c; 2294M16: a"
                     /replace="t"
     variation       148707
                     /note="242P2: a; 2294M16: g"
                     /replace="g"
     variation       151285
                     /note="242P2: a; 2294M16: g"
                     /replace="g"
     variation       151480
                     /note="242P2: t; 2294M16: c"
                     /replace="c"
     variation       152361
                     /note="242P2: a; 2294M16: g"
                     /replace="g"
     variation       152736
                     /note="242P2: t; 2294M16: c"
                     /replace="c"
     variation       152804
                     /note="242P2: a; 2294M16: c"
                     /replace="c"
     variation       152837
                     /note="242P2: c; 2294M16: t"
                     /replace="t"
     variation       154549
                     /note="242P2: a; 2294M16: g"
                     /replace="g"
     variation       155489. .155491
                     /note="242P2: atg; 2294M16: a"
                     /replace="a"
     variation       155537
                     /note="242P2: a; 2294M16: atatg"
                     /replace="atatg"
     variation       155586. .155588
                     /note="242P2: cat: 2294M16: c"
                     /replace="c"
     variation       155648. .155650
                     /note="242P2: ata; 2294M16: atgtgtgtg"
                     /replace="atgtgtgtg"
     variation       155882
                     /note="242P2: a; 2294M16: c"
                     /replace="c"
ORIGIN

Alignment Scores:
Pred. No.:              3.81e+03        Length:         163584
Score:                  44.00           Matches:        8
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            91.67%          Indels:         0
DB:                     9               Gaps:           0

US-10-725-373-5 (1-9) x AC008056 (1-163584)

Qy          1 TyrLeuSerGlyAlaCysLeuAsn 8
              ||||||||||||||||||||||||
Db     144884 TACCTTTCTGGTGCATGTTTGAAT 144907

RESULT 8
AC103486
LOCUS       AC103486               250810 bp    DNA     linear   HTG 10-MAY-2003
DEFINITION  Rattus norvegicus clone CH230-19A6, WORKING DRAFT SEQUENCE, 2
            unordered pieces.
ACCESSION   AC103486
VERSION     AC103486.6  GI:30520423
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 250810)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
```

```
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 250810)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-NOV-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 250810)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-MAY-2003) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On May 10, 2003 this sequence version replaced gi:24942663.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GHTA
                Center clone name: CH230-19A6
            -------------- Summary Statistics
                Assembly program: Atlas 3.0;
                Consensus quality: 243081 bases at least Q40
                Consensus quality: 245135 bases at least Q30
                Consensus quality: 246521 bases at least Q20
                Estimated insert size: 254009; sum-of-contigs estimation
                Quality coverage: 7x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)
            * NOTE: This sequence may represent more than one clone.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 2 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1   249098: contig of 249098 bp in length
            *   249099   249198: gap of unknown length
            *   249199   250810: contig of 1612 bp in length.
FEATURES             Location/Qualifiers
     source          1. .250810
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-19A6"
     misc_feature    1. .1599
                     /note="wgs_contig"
     misc_feature    246136. .249098
                     /note="wgs_contig"
ORIGIN

Alignment Scores:
Pred. No.:              5.58e+03        Length:         250810
Score:                  44.00           Matches:        8
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            91.67%          Indels:         0
DB:                     2               Gaps:           0

US-10-725-373-5 (1-9) x AC103486 (1-250810)

Qy          1 TyrLeuSerGlyAlaCysLeuAsn 8
              ||||||||||||||||||||||||
Db     115119 TATCTCTCAGGCGCCTGCTTGAAT 115142

RESULT 9
AC124920/c
LOCUS       AC124920               268281 bp    DNA     linear   HTG 19-SEP-2002
DEFINITION  Rattus norvegicus clone CH230-228J18, *** SEQUENCING IN PROGRESS
            ***, 2 unordered pieces.
ACCESSION   AC124920
VERSION     AC124920.3  GI:23196075
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 268281)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
```

Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G., Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P., Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M., Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W., Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K., Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J., Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M., Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A., Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A., Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C., Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J., Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J., Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J., Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A., Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E., Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E., Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S., Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L., Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K., Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C., Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L., Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R., Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F., Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J., Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H., Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D., Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J., Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C., Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K., Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J., Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F., Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K., Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V., Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O., Weinstock,G. and Gibbs,R.A.

TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 268281)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (20-JUN-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE 3 (bases 1 to 268281)
AUTHORS Rat Genome Sequencing Consortium.
TITLE Direct Submission
JOURNAL Submitted (19-SEP-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
COMMENT On Sep 19, 2002 this sequence version replaced gi:21703468.
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequening reads assembled using Atlas (http://www.hgsc.bcm.tmc.edu/projects/rat/). As a result, the sequence may extend beyond the ends of the clone and there may be contigs that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.
-------------- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
-------------- Project Information
Center project name: GYNG
Center clone name: CH230-228J18
-------------- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 235034 bases at least Q40
Consensus quality: 238312 bases at least Q30
Consensus quality: 240333 bases at least Q20
Estimated insert size: 255321; sum-of-contigs estimation
Quality coverage: 4x in Q20 bases; sum-of-contigs estimation
----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)
* NOTE: This sequence may represent more than one clone.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 265363: contig of 265363 bp in length
* 265364 265463: gap of unknown length
* 265464 268281: contig of 2818 bp in length.

FEATURES Location/Qualifiers
source 1. .268281
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-228J18"
misc_feature 1. .1201
/note="wgs_contig"
misc_feature 263541. .265363
/note="wgs_contig"
misc_feature 265464. .266507
/note="wgs_contig"

ORIGIN

Alignment Scores:

Pred. No.:	5.93e+03	Length:	268281
Score:	44.00	Matches:	8
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	91.67%	Indels:	0
DB:	2	Gaps:	0

US-10-725-373-5 (1-9) x AC124920 (1-268281)

```
Qy          1 TyrLeuSerGlyAlaCysLeuAsn 8
              ||||||||||||||||||||||||
Db      48802 TATCTCTCAGGCGCCTGCTTGAAT 48779
```

RESULT 10
AC133403/c
LOCUS AC133403 274993 bp DNA linear HTG 20-NOV-2002
DEFINITION Rattus norvegicus clone CH230-312O9, WORKING DRAFT SEQUENCE.
ACCESSION AC133403
VERSION AC133403.2 GI:25138990
KEYWORDS HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 274993)
AUTHORS Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J., Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D., Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H., Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F., Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M., Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E., Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J., Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L., Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D., Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K., Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K., Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G., Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P., Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M., Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,

```
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 274993)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-SEP-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 274993)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-NOV-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Nov 20, 2002 this sequence version replaced gi:22795083.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            ---------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            ---------- Project Information
                Center project name: KCBN
                Center clone name: CH230-31209
            ---------- Summary Statistics
                Assembly program: Phrap; version 0.990329
                Consensus quality: 263872 bases at least Q40
                Consensus quality: 266813 bases at least Q30
                Consensus quality: 268369 bases at least Q20
                Estimated insert size: 272879; sum-of-contigs estimation
                Quality coverage: 7x in Q20 bases; sum-of-contigs estimation
            ----------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)
            * NOTE: This sequence may represent more than one clone.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 1 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *       1   274993: contig of 274993 bp in length.
FEATURES            Location/Qualifiers
     source         1. .274993
                    /organism="Rattus norvegicus"
                    /mol_type="genomic DNA"
                    /db_xref="taxon:10116"
                    /clone="CH230-31209"
     misc_feature   1. .1185
                    /note="wgs_end_extension
                    clone_end:T7"
     misc_feature   2065. .2651
                    /note="clone_boundary
                    clone_end:T7
                    site:
                    end_sequence:BZ156831"
     misc_feature   complement(267606. .268445)
                    /note="clone_boundary
                    clone_end:Sp6
                    site:
                    end_sequence:BZ156848"
     misc_feature   273381. .274993
                    /note="wgs_end_extension
                    clone_end:Sp6"
ORIGIN

Alignment Scores:
Pred. No.:              6.06e+03        Length:         274993
Score:                  44.00           Matches:        8
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            91.67%          Indels:         0
DB:                     2               Gaps:           0

US-10-725-373-5 (1-9) x AC133403 (1-274993)

Qy          1 TyrLeuSerGlyAlaCysLeuAsn 8
              ||||||||||||||||||||||||
Db     152467 TATCTCTCAGGCGCCTGCTTGAAT 152444

RESULT 11
AL645535/c
LOCUS       AL645535              161363 bp    DNA     linear   ROD 09-AUG-2002
DEFINITION  Mouse DNA sequence from clone RP23-145G23 on chromosome 11,
            complete sequence.
ACCESSION   AL645535
VERSION     AL645535.16  GI:22204296
KEYWORDS    HTG.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 161363)
  AUTHORS   Bates,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (31-JUL-2002) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Aug 11, 2002 this sequence version replaced gi:22003119.
```

```
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            --------------
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. The following
            abbreviations are used to associate primary accession numbers given
            in the feature table with their source databases: Em:, EMBL; Sw:,
            SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
            database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep RP23-145G23 is
            from the RPCI-23 Mouse PAC Library
            constructed by the group of Pieter de Jong.
            For further details see http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6.
FEATURES              Location/Qualifiers
     source           1. .161363
                      /organism="Mus musculus"
                      /mol_type="genomic DNA"
                      /db_xref="taxon:10090"
                      /chromosome="11"
                      /clone="RP23-145G23"
                      /clone_lib="RPCI-23"
ORIGIN

Alignment Scores:
Pred. No.:              6.02e+03        Length:         161363
Score:                  43.00           Matches:        8
Percent Similarity:     88.89%          Conservative:   0
Best Local Similarity:  88.89%          Mismatches:     1
Query Match:            89.58%          Indels:         0
DB:                     10              Gaps:           0

US-10-725-373-5 (1-9) x AL645535 (1-161363)

Qy          1 TyrLeuSerGlyAlaCysLeuAsnLeu 9
              ||||||||||||   ||||||||||||
Db     142530 TATTTATCTGGACTCTGCTTGAACCTT 142504

RESULT 12
AC092023
LOCUS       AC092023              164991 bp    DNA     linear   PRI 29-MAR-2002
DEFINITION  Homo sapiens chromosome 3 clone RP11-13C14, complete sequence.
ACCESSION   AC092023 AC068300
VERSION     AC092023.2  GI:19807852
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 164991)
  AUTHORS   Kaul,R.K., Olson,M.V., Zhou,Y., James,R.A., Rouse,G., Wu,Z.,
            Saenphimmachak,C., Phelps,K.A., Raymond,C. and Haugen,E.D.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 164991)
  AUTHORS   Kaul,R.K., Olson,M.V., Raymond,C., Clendenning,J., Ivey,R.G. and
            Haugen,E.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-JUN-2001) Genome Center, University of Washington,
            Box 352145, Seattle, WA 98195, USA
REFERENCE   3  (bases 1 to 164991)
  AUTHORS   Kaul,R.K., Olson,M.V., Zhou,Y., James,R.A., Rouse,G., Wu,Z.,
            Saenphimmachak,C., Phelps,K.A., Raymond,C. and Haugen,E.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-MAR-2002) Genome Center, University of Washington,
            Box 352145, Seattle, WA 98195, USA
COMMENT     On Mar 29, 2002 this sequence version replaced gi:14456651.
            -------------- Genome Center
                Center: University of Washington Genome Center
                Center Code: UWGC
                Web site: http://www.genome.washington.edu
                Contact: uwgchtgs@u.washington.edu
                Drafting Center: BCM
            -------------- Project Information
                Center project name: chr-3
                Center clone name: RP11-13C14 (bc0121)
            ------------- Summary Statistics
                Sequencing vector: M13; L08821; 51% of reads
                Sequencing vector: plasmid; L08752; 49% of reads
                Chemistry: Dye-terminator ET; 45% of reads
                Chemistry: Dye-terminator Big Dye; 55% of reads
                Assembly program: Phrap; version 0.990319
                Consensus quality: 164625 bases at least Q40
                Consensus quality: 164952 bases at least Q30
                Consensus quality: 164982 bases at least Q20
                Insert size: 164991; sum-of-contigs
                Quality coverage: 7.3x in Q20 bases; sum-of-contigs
            --------------.
            Overlapping Sequences:
                5': Mapping in progress
                3': RP11-280I3 (UWGC:bc0336) AC099776, 46806-bp overlap
            --------------------
            Sequence Quality Assessment:
             This entry has been annotated with sequence quality
             estimates computed by the Phrap assembly program.
             All manually edited bases have been reduced to quality zero.
             Quality levels above 40 are expected to have less than
             1 error in 10,000 bp.
             Base-by-base quality values are not generally visible from the
             GenBank flat file format but are available as part
             of this entry's ASN.1 file.

            This sequence was finished as follows unless otherwise noted:
            all regions were either double-stranded or sequenced with an
            alternate chemistry or covered by high quality data (i.e., Phred
            quality >= 30); an attempt was made to resolve all sequencing
            problems, such as compressions and repeats; all regions were
            covered by at least one plasmid subclone or more than one M13
            subclone; and the assembly was confirmed by restriction digest.
            --------------------
            Sequence Validation:
             This sequence has been validated by Multiple Complete Digest
             fingerprinting. Comparison of the experimentally derived digest
             fragments with sequence-predicted fragments is given below.
             The electronically-digested sequence consists of both insert and
             vector, in order to accurately represent the entire circular BAC.
             Small fragments below a variable cutoff (approximately 400-800 bp)
             are not resolved in the fingerprint and hence do not appear
             in the table. There are no significant remaining discrepancies
             between the experimental and predicted values. Uniquely ordered
             fragments are separated by dashed lines.
                  HindIII                 EcoRI                   BglII
            ------------------   ------------------   ------------------
            SeqDerMap  FngrPrnt   SeqDerMap  FngrPrnt   SeqDerMap  FngrPrnt
            ---------  --------   ---------  --------   ---------  --------

            ---------              ---------              ---------
                 1583      1636        8696      8776        4390      4260
            ---------              ---------              ---------
                 6382      6557           6      <800        2067      2068
            ---------              ---------              ---------
                  512      <800          25      <800        5379      5535
```

449	<800	425	<800	1330	1285
5430	5264	901	902	5365	5316
3121	3135	688	<800	10122	10079
3145	3135	1163	1141	6473	6360
2066	2028	3545	3462	6074	6073
3615	3651	8104	8046	744	<800
179	<800	4625	4601	2983	3022
1171	1109	901	902	3025	3022
678	<800	6384	6326	6310	6360
633	<800	1265	1264	1918	1900
342	<800	1493	1472	109	<800
2338	2307	222	<800	3209	3220
1836	1765	2105	2132	497	<800
204	<800	2983	2914	32	<800
755	747	3094	3142	2110	2068
5308	5264	6997	6938	1094	1082
1639	1636	1765	1756	70	<800
3582	3651	4493	4464	2872	2879
103	<800	2165	2132	1054	1082
1641	1636	693	<800	410	<800
958	957	2284	2289	5727	5834
1263	1245	293	<800	539	<800
11457	11376	1470	1472	5297	5316
2470	2484	3384	3462	1635	1624
2311	2307	2720	2713	2989	3022
389	<800	772	<800	3129	3220
2237	2307	7202	7376	4727	4700
11358	11376	8349	8268	162	<800
408	<800	896	902	15954	16033
1247	1245	7079	6938	4805	4700
726	747	61	<800	1278	1285
3728	3651	2687	2713	91	<800
1358	1318	3628	3596	6370	6360
2993	2956	5377	5377	5892	5834
1325	1318	1671	1609	1094	1082
490	<800	4298	4251	2036	2068

530	<800	1376	1371	3762	3616
1720	1765	7485	7376	9341	9317
3247	3135	3117	3142	3564	3616
1424	1483	1720	1756	5988	6073
369	<800	5311	5164	1888	1900
356	<800	278	<800	13144	13071
1489	1483	1789	1756	2246	2212
3569	3651	20	<800	2093	2068
64	<800	5486	5377	97	<800
734	747	2108	2132	2204	2212
1067	1109	3130	3142		
964	957	2844	2914		
2879	2845	4968	4963		
1137	1109	2667	2713		
1097	1109	6378	6326		
15188	15135	3923	3891		
5306	5264	1585	1609		
471	<800	1388	1371		

Alignment Scores:

Pred. No.:	6.14e+03	Length:	164991
Score:	43.00	Matches:	7
Percent Similarity:	100.00%	Conservative:	2
Best Local Similarity:	77.78%	Mismatches:	0
Query Match:	89.58%	Indels:	0
DB:	9	Gaps:	0

US-10-725-373-5 (1-9) x AC092023 (1-164991)

```
Qy          1 TyrLeuSerGlyAlaCysLeuAsnLeu 9
              ||||||||||||:::|||:::||||||
Db     122557 TATCTGAGTGGTTCATGTATGAACCTC 122583
```

RESULT 13
AC099776/c

```
LOCUS       AC099776              171347 bp    DNA     linear   PRI 14-FEB-2002
DEFINITION  Homo sapiens chromosome 3 clone RP11-280I3, complete sequence.
ACCESSION   AC099776 AC074280
VERSION     AC099776.2  GI:18657048
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 171347)
  AUTHORS   Kaul,R.K., Olson,M.V., Zhou,Y., James,R.A., Rouse,G., Wu,Z.,
            Saenphimmachak,C., Phelps,K.A., Raymond,C. and Haugen,E.D.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 171347)
  AUTHORS   Kaul,R.K., Olson,M.V., Raymond,C. and Haugen,E.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-NOV-2001) Genome Center, University of Washington,
            Box 352145, Seattle, WA 98195, USA
REFERENCE   3  (bases 1 to 171347)
  AUTHORS   Kaul,R.K., Olson,M.V., Zhou,Y., James,R.A., Rouse,G., Wu,Z.,
```

Saenphimmachak,C., Phelps,K.A., Raymond,C. and Haugen,E.D.
TITLE Direct Submission
JOURNAL Submitted (14-FEB-2002) Genome Center, University of Washington, Box 352145, Seattle, WA 98195, USA
COMMENT On Feb 14, 2002 this sequence version replaced gi:17027291.
---------- Genome Center
Center: University of Washington Genome Center
Center Code: UWGC
Web site: http://www.genome.washington.edu
Contact: uwgchtgs@u.washington.edu
Drafting Center: BCM
---------- Project Information
Center project name: chr-3
Center clone name: RP11-280I3 (bc0336)
---------- Summary Statistics
Sequencing vector: plasmid; 44% of reads
Sequencing vector: M13; L08821; 56% of reads
Sequencing vector: plasmid; L08752; 0% of reads
Chemistry: Dye-terminator ET; 39% of reads
Chemistry: Dye-primer Bodipy; 4% of reads
Chemistry: Dye-terminator Big Dye; 57% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 171170 bases at least Q40
Consensus quality: 171330 bases at least Q30
Consensus quality: 171347 bases at least Q20
Insert size: 171346; sum-of-contigs
Quality coverage: 8.0x in Q20 bases; sum-of-contigs
----------.
Overlapping Sequences:
5': RP11-13C14 (UWGC:bc0121) AC092023
3': RP11-435M24 (UWGC:bc0420) AC092420, 3064-bp overlap

Sequence Quality Assessment:
This entry has been annotated with sequence quality estimates computed by the Phrap assembly program. All manually edited bases have been reduced to quality zero. Quality levels above 40 are expected to have less than 1 error in 10,000 bp. Base-by-base quality values are not generally visible from the GenBank flat file format but are available as part of this entry's ASN.1 file.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., Phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest.

Sequence Validation:
This sequence has been validated by Multiple Complete Digest fingerprinting. Comparison of the experimentally derived digest fragments with sequence-predicted fragments is given below. The electronically-digested sequence consists of both insert and vector, in order to accurately represent the entire circular BAC. Small fragments below a variable cutoff (approximately 400-800 bp) are not resolved in the fingerprint and hence do not appear in the table. There are no significant remaining discrepancies between the experimental and predicted values. Uniquely ordered fragments are separated by dashed lines.

HindIII		BglII		EcoRI	
SeqDerMap	FngrPrnt	SeqDerMap	FngrPrnt	SeqDerMap	FngrPrnt
1748	1716	5104	4922	8696	8914
6382	6590	2067	2079	6	<800
512	<800	4720	4922	1720	1746
449	<800	3762	3779	5311	5156
5459	5330	9341	9234	278	<800
471	<800	3566	3583	1789	1746
4419	4423	5988	6069	20	<800
376	<800	1888	1875	5487	5395
2049	2022	13116	12878	2109	2110
7643	7617	2246	2250	3130	3320
1987	2022	2091	2079	2844	2875
5893	6014	97	<800	4968	5156
6633	6590	2204	2250	2663	2704
6647	6590	8350	8230	6354	6301
1828	1884	7183	7228	3923	3873
3141	3131	2563	2631	1583	1637
4271	4244	3291	3346	1388	1375
7989	7826	4260	4231	1517	1490
1083	1077	1506	1447	1683	1637
3123	3131	2415	2453	2529	2567
1362	1328	2966	2985	567	<800
5413	5330	6959	6931	1127	1111
1890	1884	747	757	452	<800
3627	3607	1624	1594	1516	1490
55	<800	4850	4922	4657	4576
2336	2319	1663	1594	125	<800
6686	6590	7719	7669	11257	11227
1886	1884	281	<800	5162	5156
2442	2454	3112	3171	574	<800
254	<800	8712	8596	562	<800
7392	7364	488	<800	8296	8219
224	<800	760	757	3316	3423
476	<800	2042	2079	746	761
5240	5150	2396	2453	426	<800
334	<800	20802	21047	9062	8914
4224	4244	3471	3475	4685	4576
2487	2454	1318	1276	3832	3873
2604	2654	2049	2079	3973	3873
5186	5150	3761	3779	5273	5156
1256	1227	4354	4400	1538	1490

161	<800	9087	8991	2211	2201
4391	4423	1330	1276	3495	3423
1892	1884	2193	2250	1680	1637
541	<800	1604	1594	8426	8219
3674	3607			608	<800
2525	2562			140	<800
5552	5489			3404	3423
6957	6899			3010	3025
873	886			1403	1375
6052	6014			5673	5552
347	<800			2391	2436
6135	6014			4956	4919
511	<800			12922	13091
3873	3821			1522	1490
487	<800			3061	3174
4085	4063				

Alignment Scores:

Pred. No.:	6.35e+03	Length:	171347
Score:	43.00	Matches:	7
Percent Similarity:	100.00%	Conservative:	2
Best Local Similarity:	77.78%	Mismatches:	0
Query Match:	89.58%	Indels:	0
DB:	9	Gaps:	0

US-10-725-373-5 (1-9) x AC099776 (1-171347)

```
Qy          1 TyrLeuSerGlyAlaCysLeuAsnLeu 9
              |||||||||||:::|||:::||||||
Db     166976 TATCTGAGTGGTTCATGTATGAACCTC 166950
```

```
RESULT 14
AC010247/c
LOCUS       AC010247               95663 bp    DNA     linear   PRI 13-JUL-2002
DEFINITION  Homo sapiens chromosome 19 clone CTC-378H22, complete sequence.
ACCESSION   AC010247
VERSION     AC010247.9  GI:21743752
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 95663)
  AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 95663)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-SEP-1999) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE   3  (bases 1 to 95663)
  AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-JUL-2002) DOE Joint Genome Institute, 2800 Mitchell
            Drive, Walnut Creek, CA 94598, USA
REFERENCE   4  (bases 1 to 95663)
  AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-JUL-2002) DOE Joint Genome Institute, 2800 Mitchell
            Drive, Walnut Creek, CA 94598, USA
COMMENT     On Jul 13, 2002 this sequence version replaced gi:21637454.
            Draft Sequence Produced by DOE Joint Genome Institute
            www.jgi.doe.gov
            Finishing Completed at Stanford Human Genome Center
            www-shgc.stanford.edu
            Quality: Phrap Quality >=40 99.9% of Sequence;
            Estimated Total Number of Errors is 0.1.
            NOTE:  Small insert shatter library only 85964-86181.  226bp single
            subclone 9479-9697.
FEATURES             Location/Qualifiers
     source          1. .95663
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="19"
                     /clone="CTC-378H22"
     misc_feature    join(9479. .9697,85964. .86181)
                     /note="NOTE:  Small insert shatter library only
                     85964-86181. 226bp single subclone 9479-9697."
ORIGIN
```

Alignment Scores:

Pred. No.:	6.04e+03	Length:	95663
Score:	42.00	Matches:	7
Percent Similarity:	100.00%	Conservative:	2
Best Local Similarity:	77.78%	Mismatches:	0
Query Match:	87.50%	Indels:	0
DB:	9	Gaps:	0

US-10-725-373-5 (1-9) x AC010247 (1-95663)

```
Qy          1 TyrLeuSerGlyAlaCysLeuAsnLeu 9
              :::|||||||||||||||:::||||||
Db      15673 TTTCTGAGTGGTGCTTGTATCAATTTA 15647
```

```
RESULT 15
AC122164
LOCUS       AC122164               112427 bp    DNA     linear   PLN 19-FEB-2003
DEFINITION  Medicago truncatula clone mth2-24h4, complete sequence.
ACCESSION   AC122164
VERSION     AC122164.5  GI:24414295
KEYWORDS    HTG.
SOURCE      Medicago truncatula (barrel medic)
  ORGANISM  Medicago truncatula
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
            Medicago.
REFERENCE   1  (bases 1 to 112427)
  AUTHORS   Shaull,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B.,
            Cook,D., Kim,D. and Roe,B.A.
  TITLE     Medicago truncatula BAC Clone mth2-24h4
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 112427)
  AUTHORS   Shaull,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B.,
            Cook,D., Kim,D. and Roe,B.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-MAY-2002) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
REFERENCE   3  (bases 1 to 112427)
  AUTHORS   Shaull,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B.,
            Cook,D., Kim,D. and Roe,B.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-JUL-2002) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
REFERENCE   4  (bases 1 to 112427)
```

```
  AUTHORS   Shaull,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B.,
            Cook,D., Kim,D. and Roe,B.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-AUG-2002) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
REFERENCE   5  (bases 1 to 112427)
  AUTHORS   Shaull,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B.,
            Cook,D., Kim,D. and Roe,B.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-OCT-2002) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
REFERENCE   6  (bases 1 to 112427)
  AUTHORS   Shaull,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B.,
            Cook,D., Kim,D. and Roe,B.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-FEB-2003) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
REFERENCE   7  (bases 1 to 112427)
  AUTHORS   Shaull,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B.,
            Cook,D., Kim,D. and Roe,B.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-FEB-2003) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
COMMENT     On Oct 26, 2002 this sequence version replaced gi:21955031.
            -------------- Genome Center
            Center: Department Of Chemistry And Biochemistry
            The University Of Oklahoma
            Center code:UOKNOR
            ---------------------.
FEATURES             Location/Qualifiers
     source          1. .112427
                     /organism="Medicago truncatula"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:3880"
                     /clone="mth2-24h4"
                     /clone_lib="Medicago truncatula BAC library H2"
ORIGIN

Alignment Scores:
Pred. No.:              6.97e+03        Length:         112427
Score:                  42.00           Matches:        7
Percent Similarity:     100.00%         Conservative:   1
Best Local Similarity:  87.50%          Mismatches:     0
Query Match:            87.50%          Indels:         0
DB:                     8               Gaps:           0

US-10-725-373-5 (1-9) x AC122164 (1-112427)

Qy          1 TyrLeuSerGlyAlaCysLeuAsn 8
              |||:::|||||||||||||||||||
Db      68972 TATATGTCAGGTGCATGCCTGAAT 68995


Search completed: May 17, 2005, 19:15:17
Job time : 1440.5 secs
```